

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ MÔN KHO DỮ LIỆU VÀ OLAP**  
**Phân tích dữ liệu về nạn xả súng ở**  
**Mỹ từ 2013 – 2015**

**Sinh viên thực hiện:** 19521783- Lê Văn Long  
19520854- Phạm Hoàng Phúc  
**Lớp:** IS217.M22.HTCL  
**Giảng viên hướng dẫn:** TS Đỗ Thị Minh Phụng

**PHỤ LỤC**

<b>CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI.....</b>	<b>4</b>
1. Lý do chọn đề tài .....	4
2. Giới thiệu về Dataset .....	4
<b>CHƯƠNG 2: QUÁ TRÌNH TẠO PROJECT VÀ THAO TÁC DỮ LIỆU .....</b>	<b>8</b>
1. Quá trình SSIS: .....	8
2. Quá trình SSAS: .....	17
3. Tiến hành truy vấn 10 câu query trên Cube, MDX và Pivot Excel.....	25
4. Quá trình tạo Report (SSRS) .....	38
<b>CHƯƠNG 3: DATA MINING.....</b>	<b>74</b>
1. Giới thiệu Dataset .....	74
2. Quá trình Data Mining.....	76
3. Tiến hành phân tích. ....	84

## CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

### 1. Lý do chọn đề tài

- Hiện nay, nạn xả súng tại Mỹ đang là một vấn nạn lớn, gây ra thiệt hại nghiêm trọng về tính mạng con người. Nguyên nhân xả súng đến từ nhiều lý do, trong đó đáng kể nhất là sự quản lý lỏng lẻo của chính quyền, nạn phân biệt chủng tộc tôn giáo, bắt nạt, tư thù cá nhân, mục đích chính trị,...
- Để hiểu rõ hơn về những lí do trên, nhóm chọn đề tài “Phân tích dữ liệu về nạn xả súng ở Mỹ từ 2013 – 2015”

### 2. Giới thiệu về Dataset

#### 2.1 Giới thiệu về nguồn dữ liệu

##### 2.1.1 Tên dataset

- Tên dataset: Gun Violence Data

##### 2.1.2 Đơn vị cung cấp

- Đây là bộ dữ liệu mô tả về hơn 260,000 vụ xả súng ở Mỹ từ năm 2013 tới 2015, được cung cấp bởi tài khoản Kaggle có tên James Ko. Hiện không có thông tin chi tiết về tác giả, nhưng dataset có 32,500 lượt download, 226,000 lượt xem và 858 lượt upvote cho thấy đây là một dataset có độ chính xác cao.

##### 2.1.3 Mô tả số dòng, cột, thời gian thu thập

- Dataset gồm 29 cột và 239,675 dòng
- Data được thu thập từ 2008 tới 2013

##### 2.1.4 Nguồn tải dataset

- Link dataset: <https://www.kaggle.com/jameslko/gun-violence-data>

##### 2.1.5 Định hướng chủ đề của dataset

- Dataset định hướng theo chủ đề: Social science (khoa học xã hội)

##### 2.1.6 Mô tả chi tiết các cột thuộc tính dataset

Sau khi trích lọc các cột và dòng không cần thiết cho mục đích của đồ án, dataset còn 21 cột và 140179 dòng.

STT	Tên thuộc tính	Ý nghĩa	Kiểu dữ liệu
1	incident_id	Mã sự kiện	String
2	date	Ngày sự kiện	Date
3	state	Tiểu bang xảy ra sự kiện	String
4	city or country	Thành phố hoặc hạt xảy ra sự kiện	String
5	address	Địa chỉ xảy ra sự kiện	String
6	n_killed	Số trường hợp thiệt mạng	Int
7	n_injured	Số trường hợp bị thương	Int
8	congressional_district	Khu vực quốc hội	Int
9	gun_stolen	Súng bị trộm	Int
10	gun_type	Loại súng	Int
11	incident_characteristics	Hành động của mỗi cá thể liên quan	String
12	n_guns_involved	Số lượng súng trong sự kiện	Int
13	participant_age	Tuổi của các cá thể liên quan	String
14	participant_age_group	Độ tuổi của các cá thể liên quan	String
15	participant_gender	Giới tính của các cá thể liên quan	String
16	participant_relationship	Quan hệ giữa các cá thể liên quan	String
17	participant_status	Trạng thái của các cá thể sau sự kiện	String

18	participant_type	Vai trò của các cá thể trong sự kiện	String
19	state_house_district	Khu tiểu bang quốc hội	Int
20	state_senate_district	Khu tiểu bang nghị viện	Int

## 2.2 Các câu truy vấn:

- Câu 1: Số vụ xả súng tại ban Nevada qua từng năm:
- Câu 2: Tìm ra ngày chết chóc nhất của nước Mỹ năm 2014:
- Câu 3: Theo từng tháng, năm liệt kê số vụ xả súng ở từng bang
- Câu 4: Tên tiểu bang có số lượng người tử vong >500 trong các vụ xả súng Năm 2014
- Câu 5: với mỗi tiểu ban, đưa ra 3 thành phố có số vụ xả súng nhiều nhất
- Câu 6: Tìm ra top 3 tiểu bang an toàn nhất nước Mỹ trong năm 2014
- Câu 7: Tìm ra các thành phố có số vụ xả súng >10 và <20
- Câu 8: Liệt kê số vụ xả súng của các tiểu bang nằm ngoài DC
- Câu 9: Tính tỉ lệ % vụ xả súng có nữ liên quan.
- Câu 10: Thống kê số vụ xả súng theo quan hệ của nạn nhân và hung thủ.
- Câu 11: Tỉ lệ giữa nam nữ trong các vụ xả súng
- Câu 12 trung bình 1 ngày có bao nhiêu người chết và bị thương
- Câu 13. thống kê top 3 các nguyên nhân gây xả súng hàng đầu
- Câu 14: bang nào có nhiều người chết vì xả súng nhất và ít nhất
- Câu 15: Thống kê số vụ xả súng ở West coast

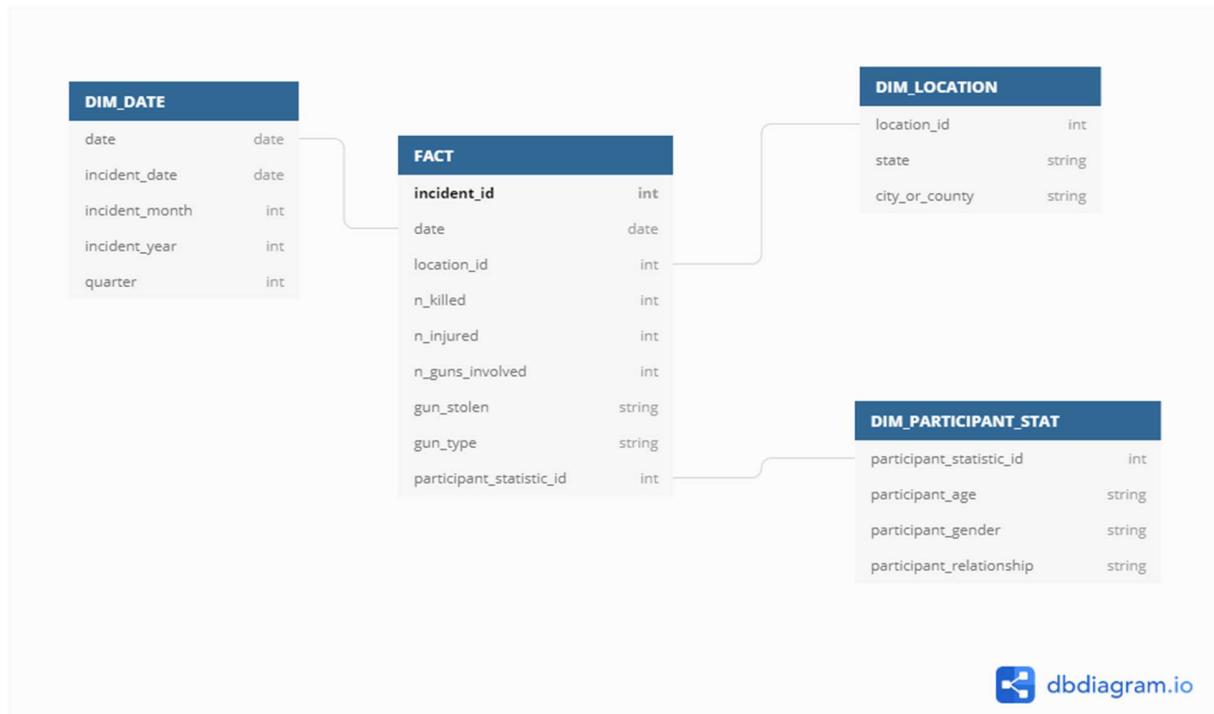


## CHƯƠNG 2: QUÁ TRÌNH TẠO PROJECT VÀ THAO TÁC DỮ LIỆU

### 1. Quá trình SSIS:

#### 1.1 Thiết kế kho dữ liệu

##### 1.1.1 Lượt đồ kho dữ liệu (lượt đồ bông tuyết)



##### 1.1.2 Mô tả chi tiết bảng FACT

STT	Tên thuộc tính	Ý nghĩa	Ràng buộc	Kiểu dữ liệu
1	incident_id	Mã sự kiện	PK	Int
2	date	Ngày sự xảy ra sự kiện	FK	Int
3	location_id	Mã vị trí	FK	Int
4	n_killed	Số tử vong		Int
5	n_injured	Số bị thương		Int
6	n_guns_involved	Số lượng súng		Int

7	gun_stolen	Súng bị trộm		String
8	gun_type	Loại súng		String
9	participant_statistic_id	Mã số liệu của các cá thể liên quan	FK	Int

### 1.1.3 Mô tả chi tiết các bảng dimension

#### 1.1.3.1 DIM\_DATE

STT	Tên thuộc tính	Ý nghĩa	Ràng buộc	Kiểu dữ liệu
1	date	ngày xảy ra sự kiện	PK	Date
2	incident_month	Tháng xảy ra sự kiện		Int
3	incident_year	Năm xảy ra sự kiện		Int
4	incident_quarter	Quý xảy ra sự kiện		Int

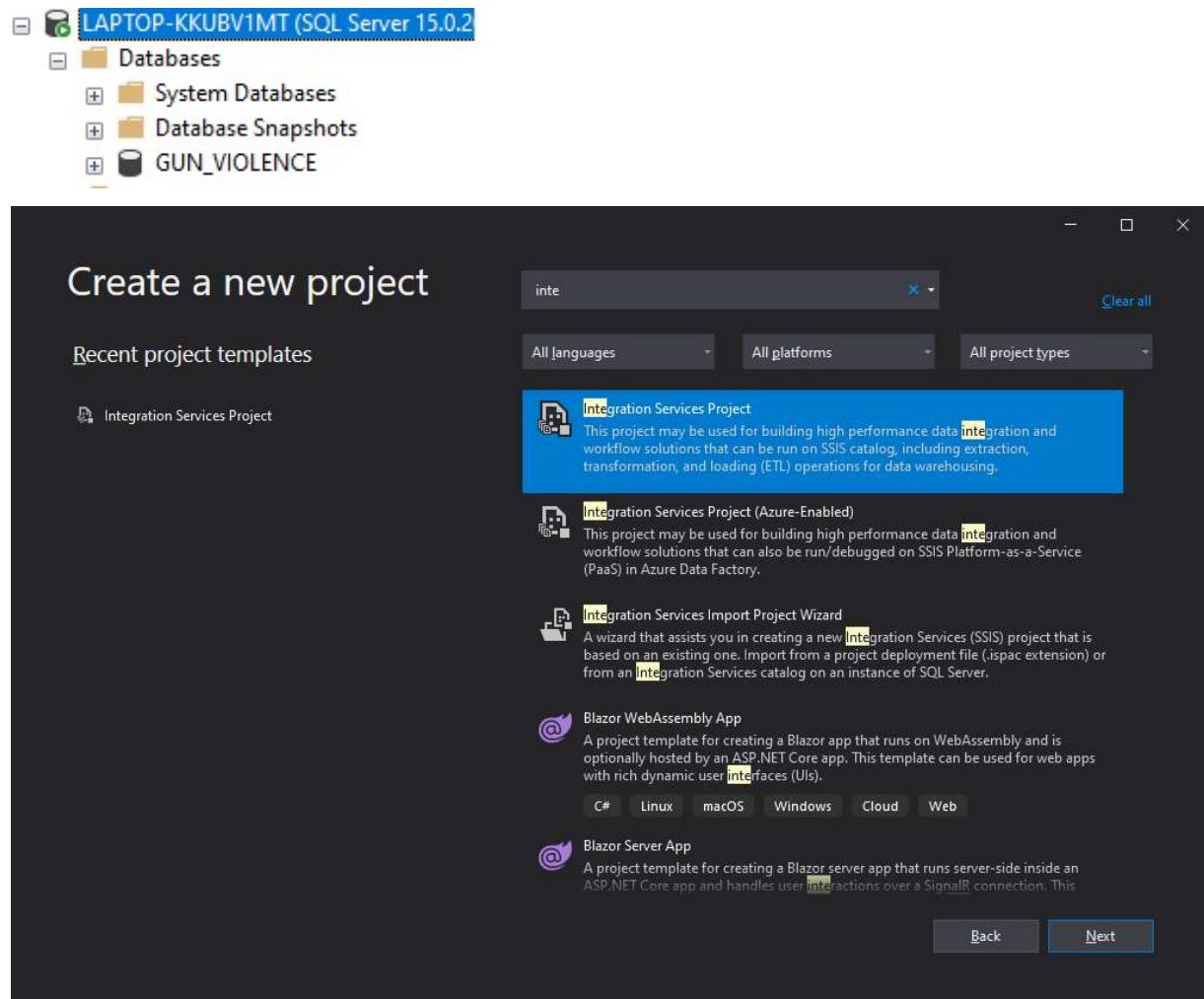
#### 1.1.3.2 DIM\_PARTICIPANT\_STAT

STT	Tên thuộc tính	Ý nghĩa	Ràng buộc	Kiểu dữ liệu
1	participant_statistic_id	Mã số liệu về người liên quan	PK	Int
2	participant_age	Độ tuổi của người liên quan		String
3	male	Số lượng nam trong vụ án		int
4	female	Số lượng nữ trong vụ án		int

## 2. Thực hiện SSIS trên Visual Studio

Bước 1: Clean data, xóa bỏ data lỗi và thùa bằng Excel

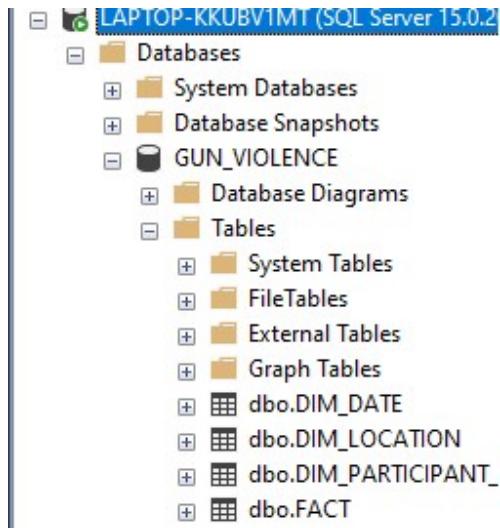
Bước 2: Tạo Project Integration Services Projects trong Visual Studio và database rỗng trên SQL Server



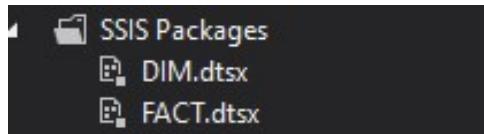
Bước 3: Tiến hành chia dữ liệu trong Excel ra bảng Fact và các bảng Dim

	DIM_DATE.xls	4/1/2022 9:21 AM	Microsoft Excel 97...
	DIM_LOCATION.xls	3/31/2022 8:21 PM	Microsoft Excel 97... 2,280 KB
	DIM_PARTICIPANT_STAT.xls	4/1/2022 8:26 AM	Microsoft Excel 97... 4,295 KB
	FACT.xls	3/31/2022 8:13 PM	Microsoft Excel 97... 6,797 KB
	location.xlsx	3/31/2022 4:32 PM	Microsoft Excel W... 184 KB

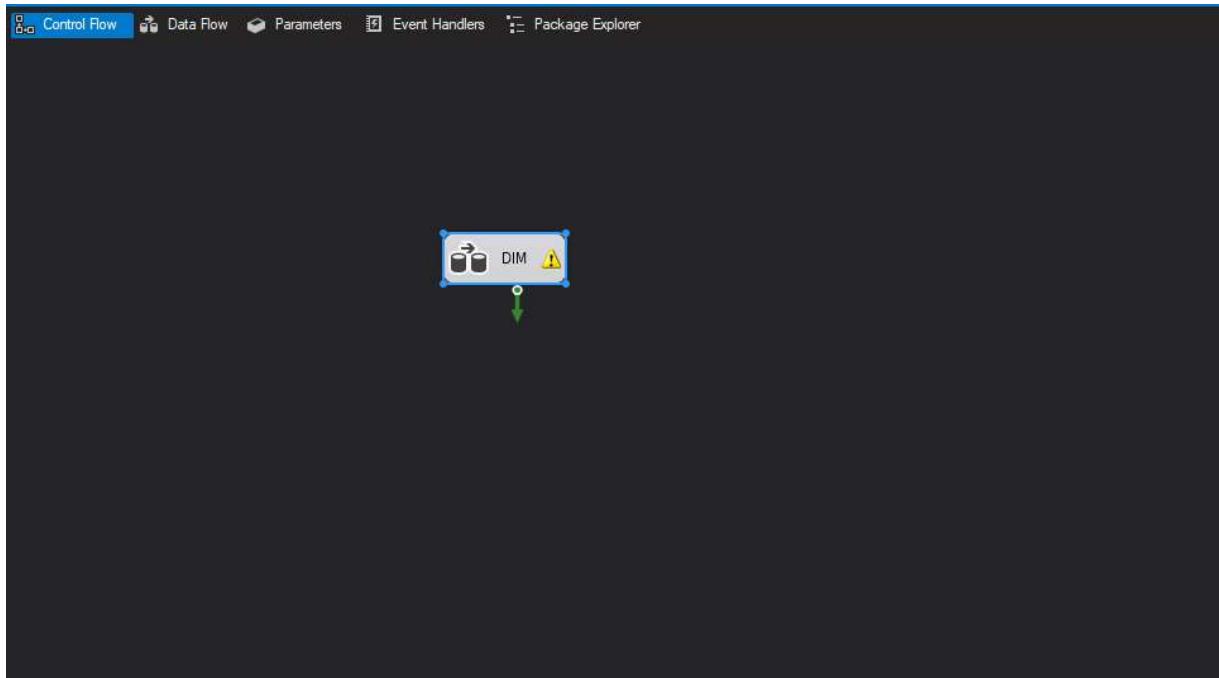
Bước 4: Tạo các bảng trống trong Database đã tạo với cấu trúc giống bước 3



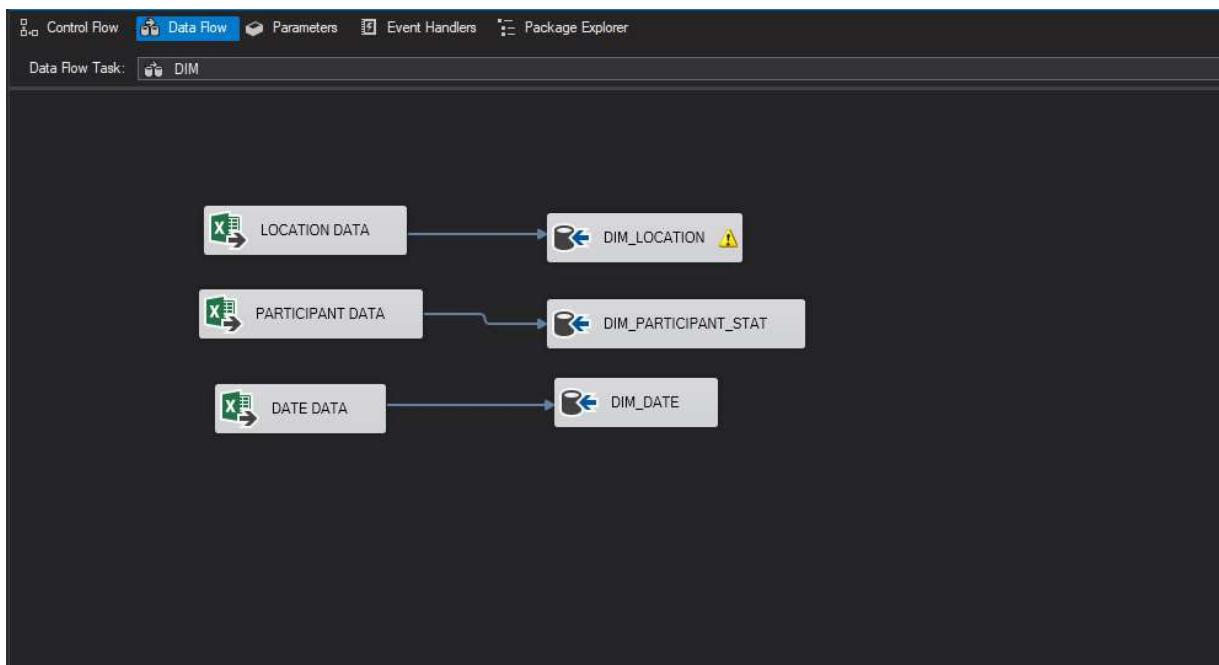
Bước 5: Tạo 2 package trong Visual Studio để chứa các thao tác đỗ data vào các bảng Dim và bảng Fact.



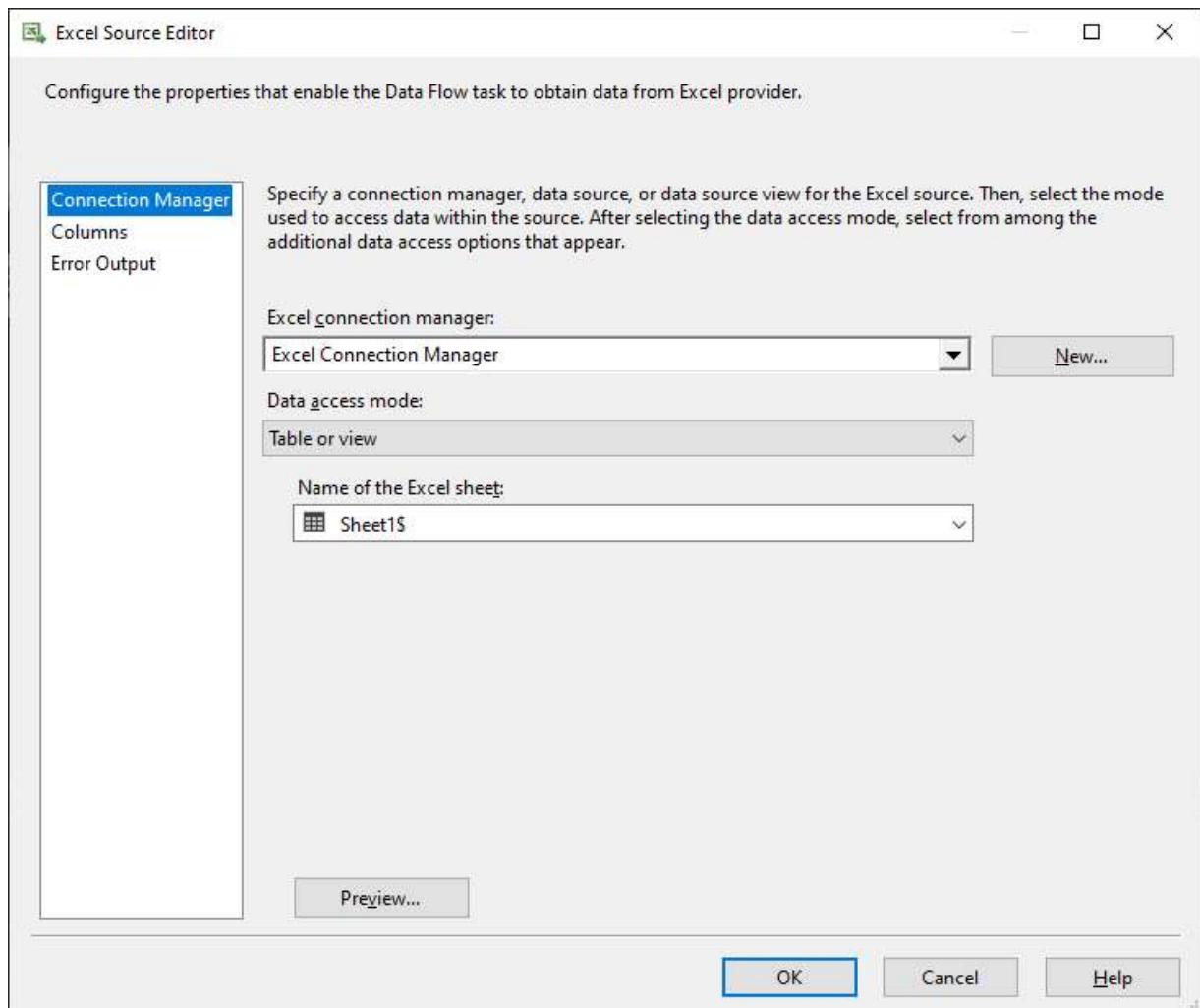
Bước 6: Tại Packet Dim, thiết lập các Data Source và Destination



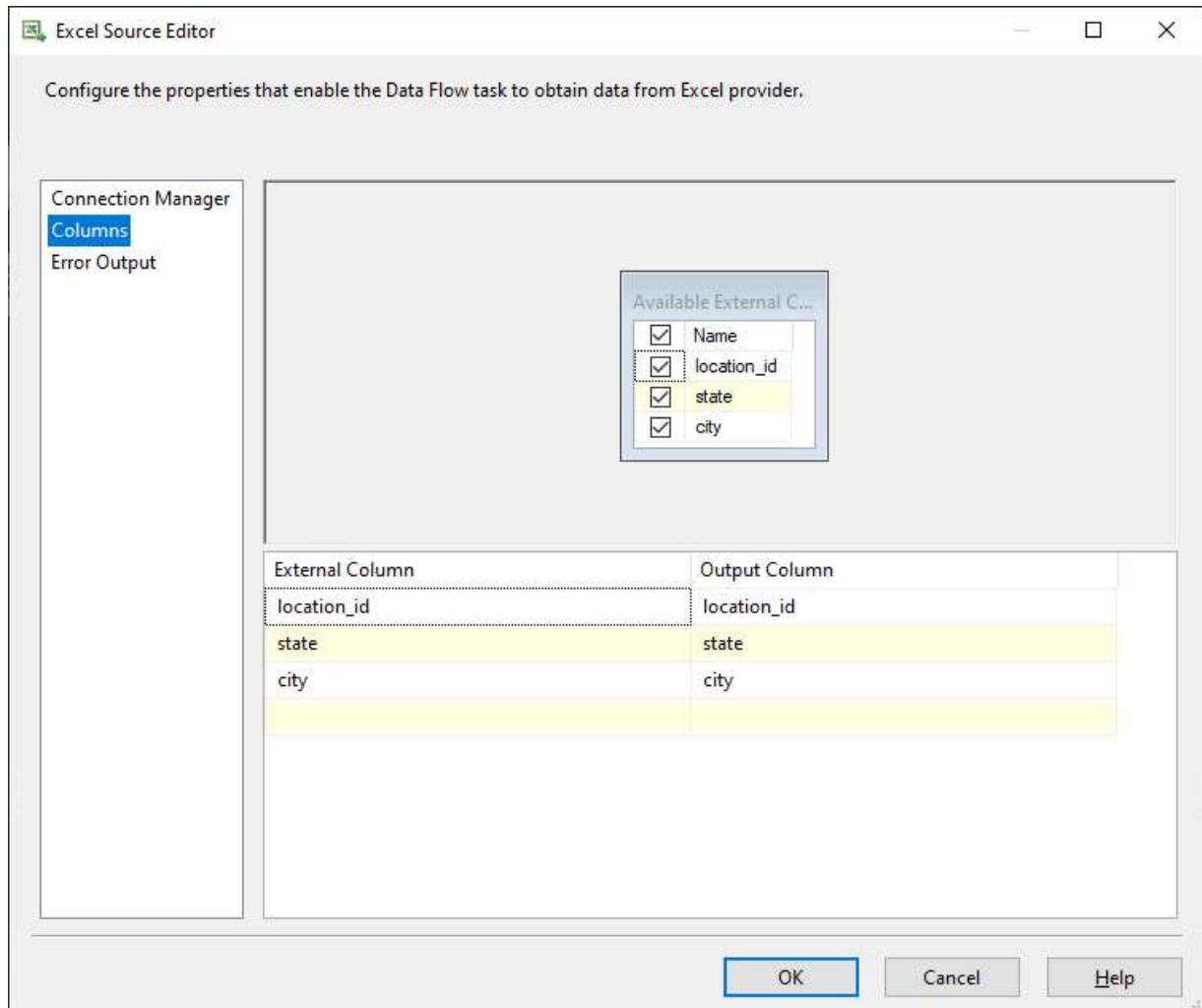
Tạo Data Flow Task



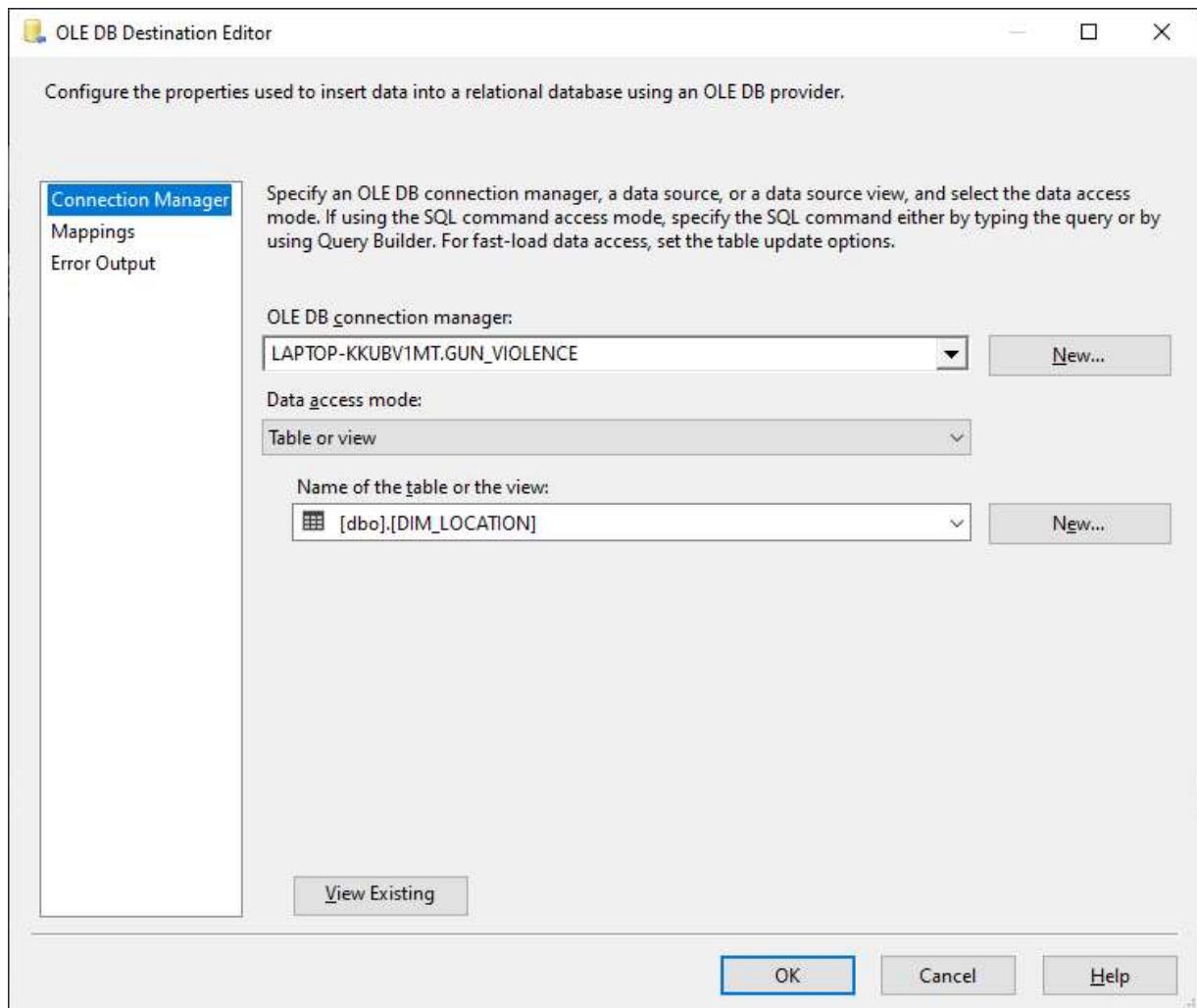
Thiết lập Data Source và Destination



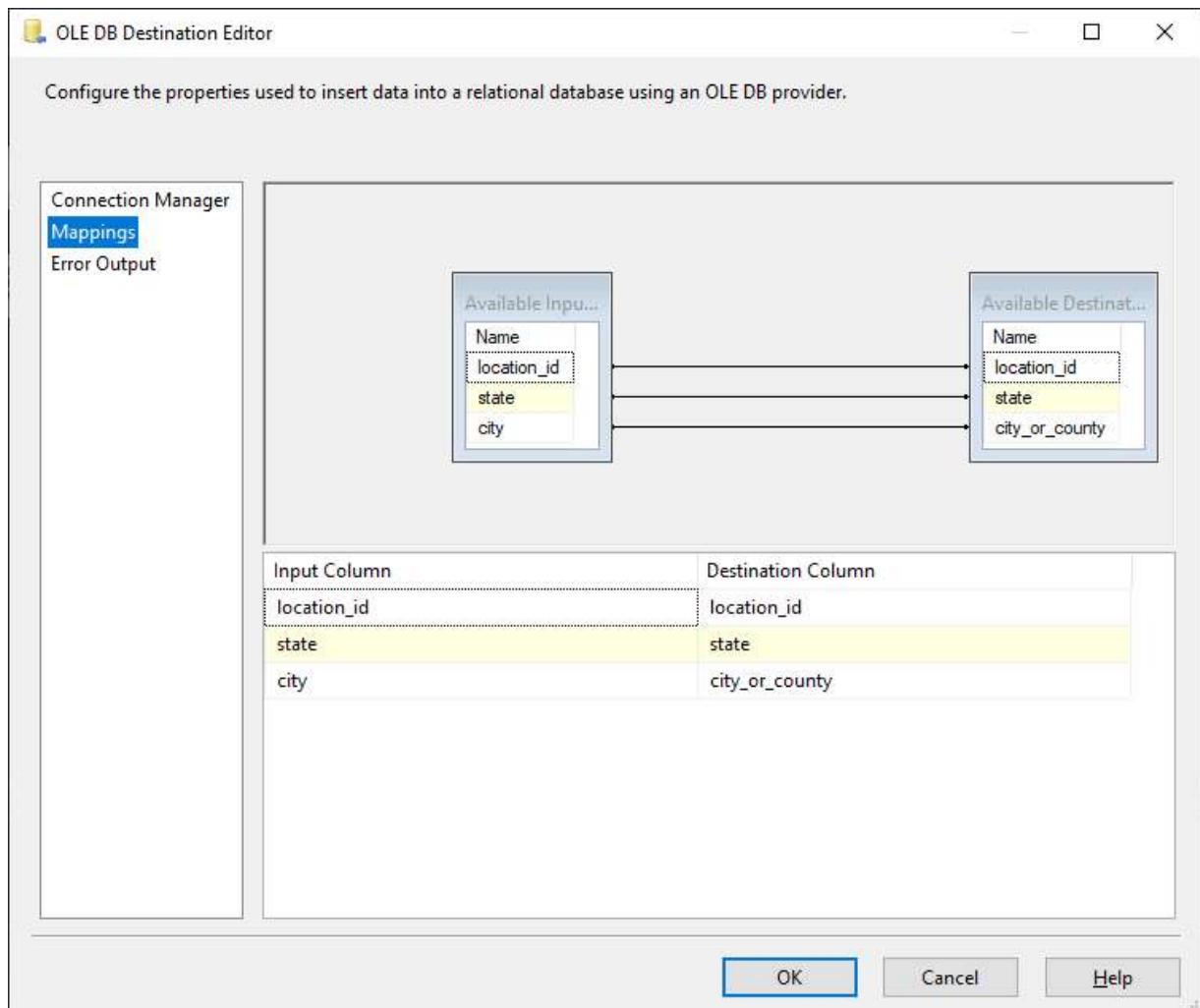
Ở mỗi Data source, browse tới file excel



Chỉ giữ lại các cột cần dùng

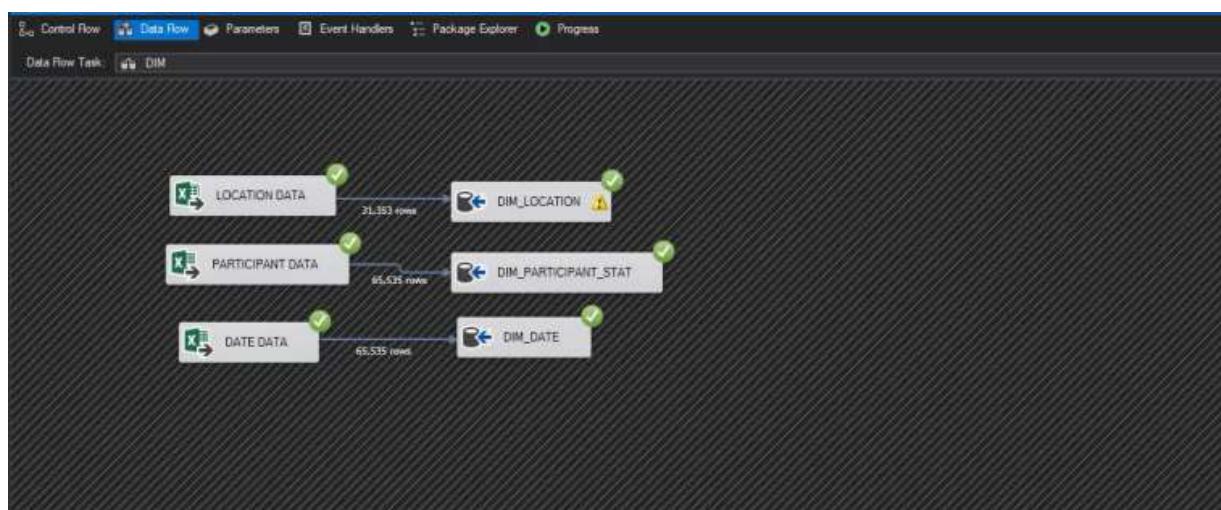


Ở mỗi Data Souce thì set up tên server và bảng sẽ đỗ data vào như hình trên



Map các cột từ file Excel với các cột trong database

Bước 7 Run Project để đổ data từ file excel vào database



Bước 8: Làm tương tự cho Fact packet

Bước 9: Chạy query để tạo khóa chính và khóa ngoại

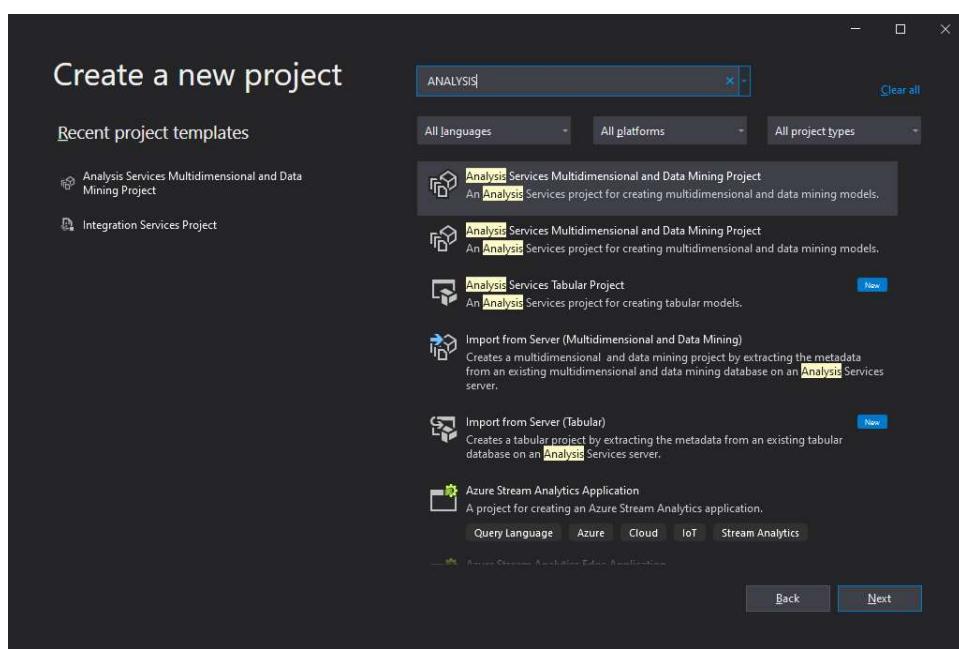
```
ALTER TABLE FACT
ADD FOREIGN KEY (date) REFERENCES DIM_DATE(date);
ALTER TABLE FACT
ADD FOREIGN KEY (location_id) REFERENCES DIM_LOCATION(location_id);
ALTER TABLE FACT
ADD FOREIGN KEY (participant_stat_id) REFERENCES DIM_PARTICIPANT_STAT(participant_stat_id);
```

Link sản phẩm:

<https://drive.google.com/drive/folders/1io7yIEiK61xupEZkSw56CyEH0E1REenL?usp=sharing>

## 2. Quá trình SSAS:

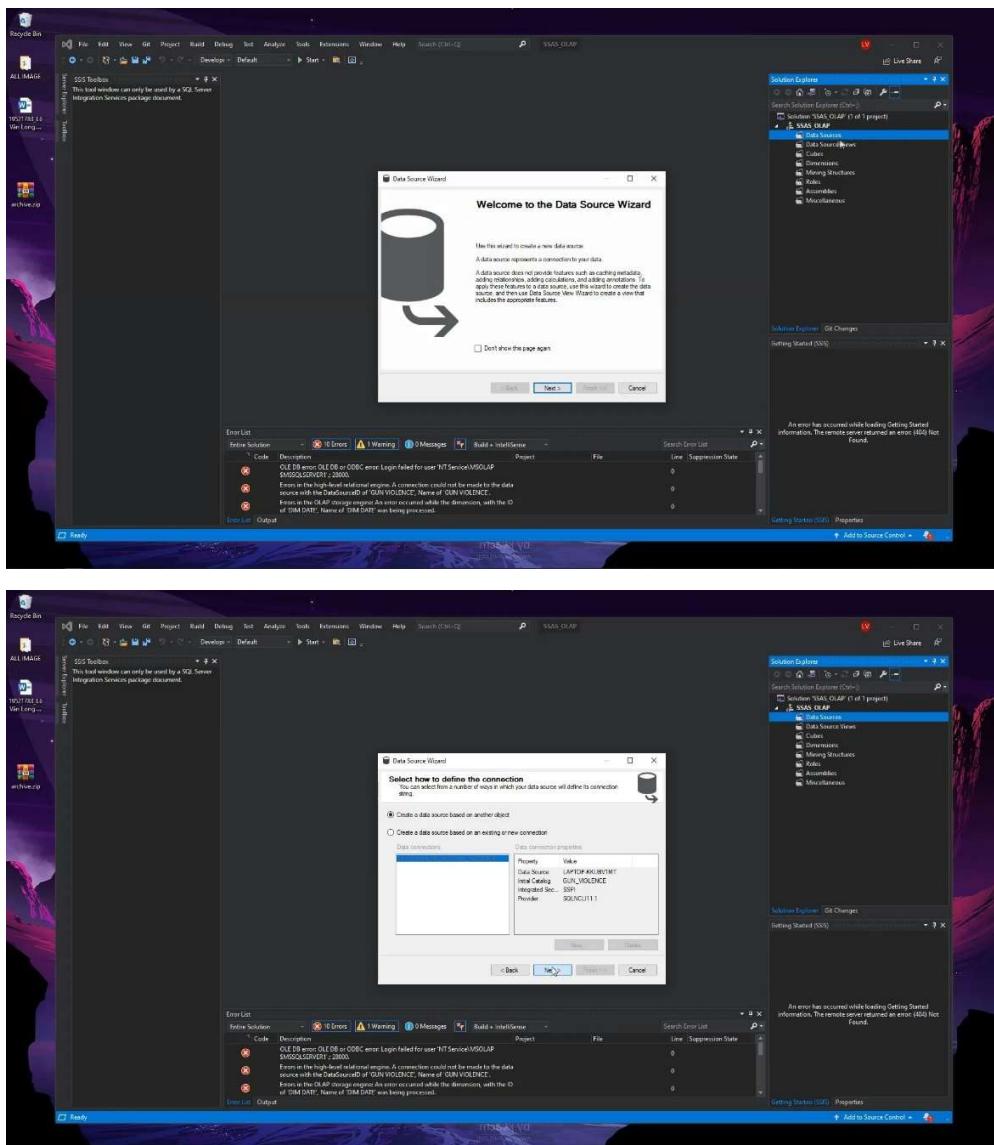
Bước 1: Chọn File /New Project / Chọn Analysis Services

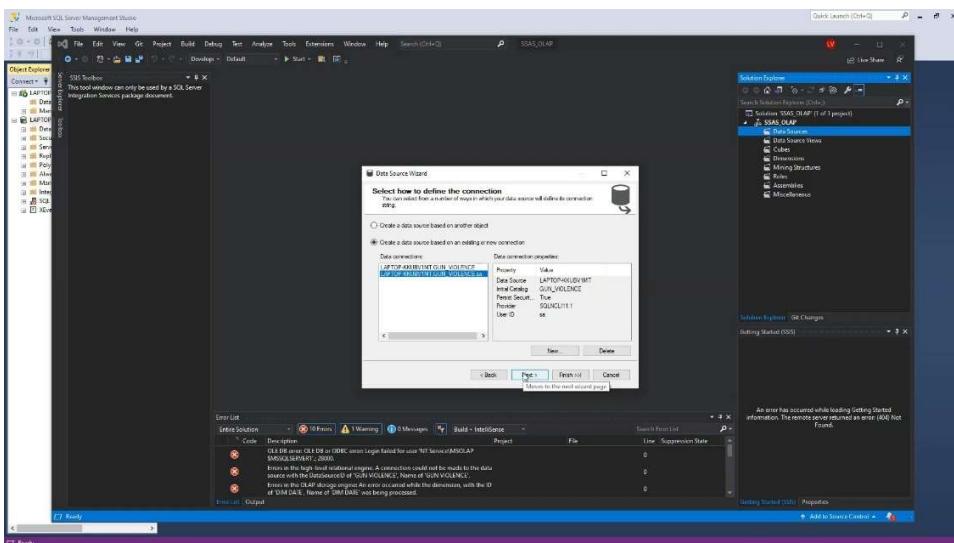
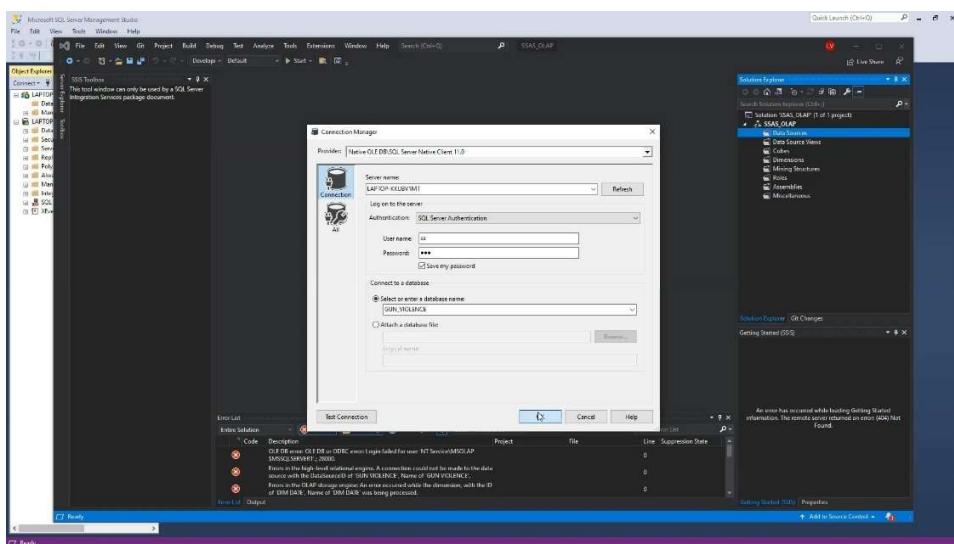


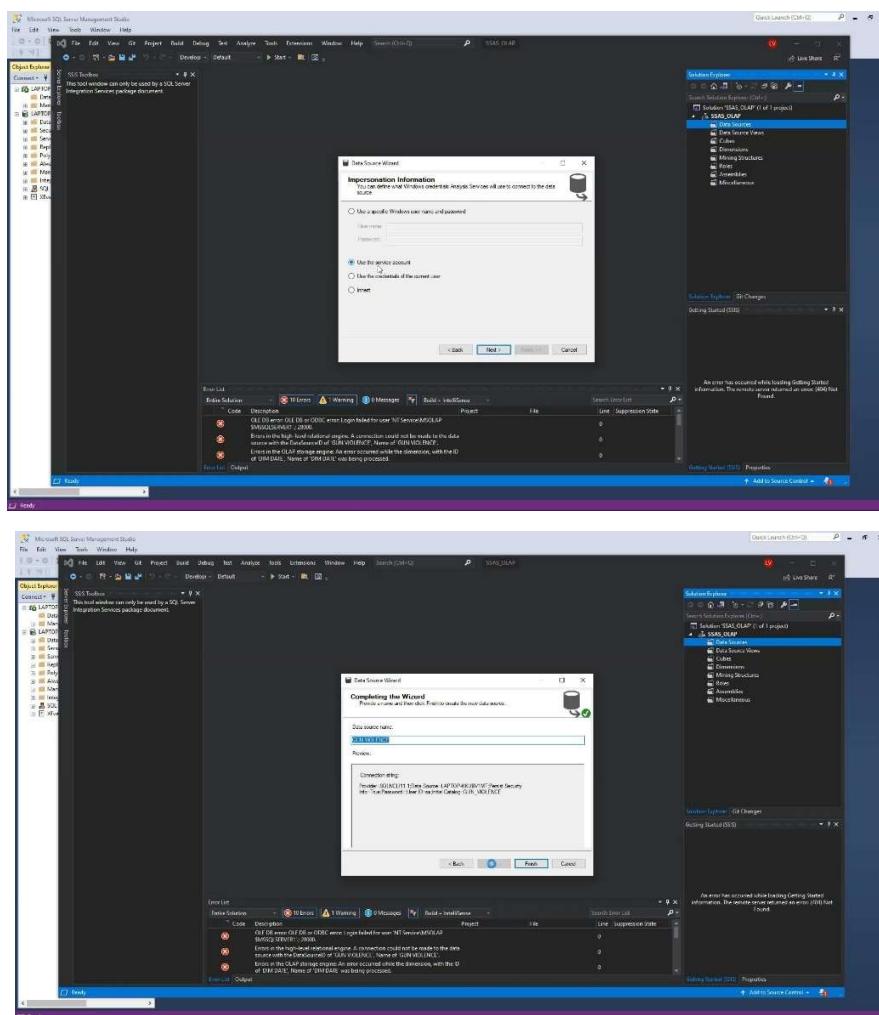
Bước 2: Xác lập dữ liệu nguồn (Data Source)

## KHO DỮ LIỆU VÀ OLAP

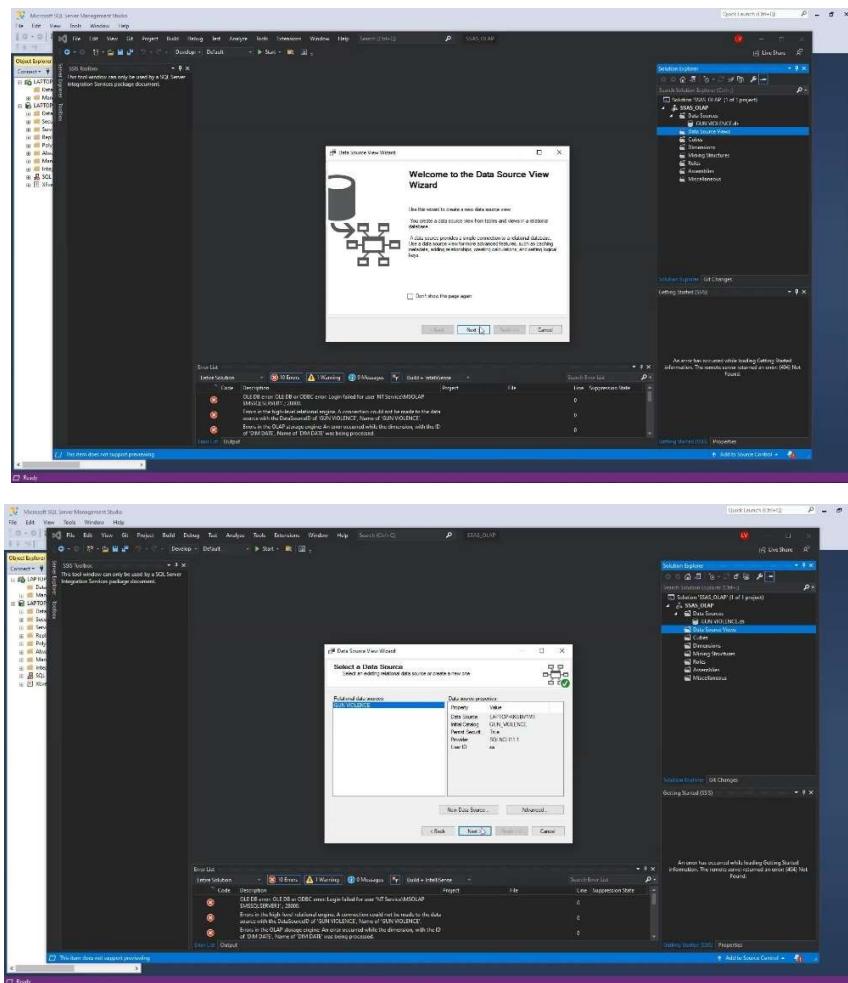
IS217.M22.HTCL

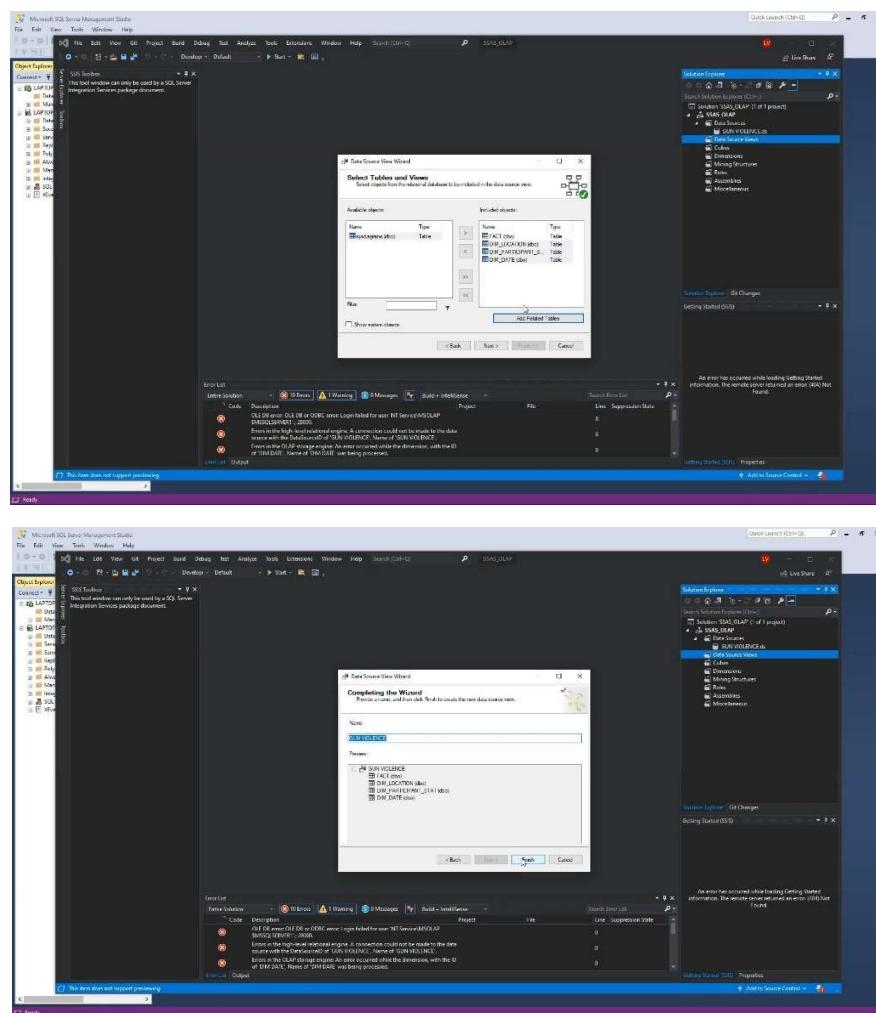




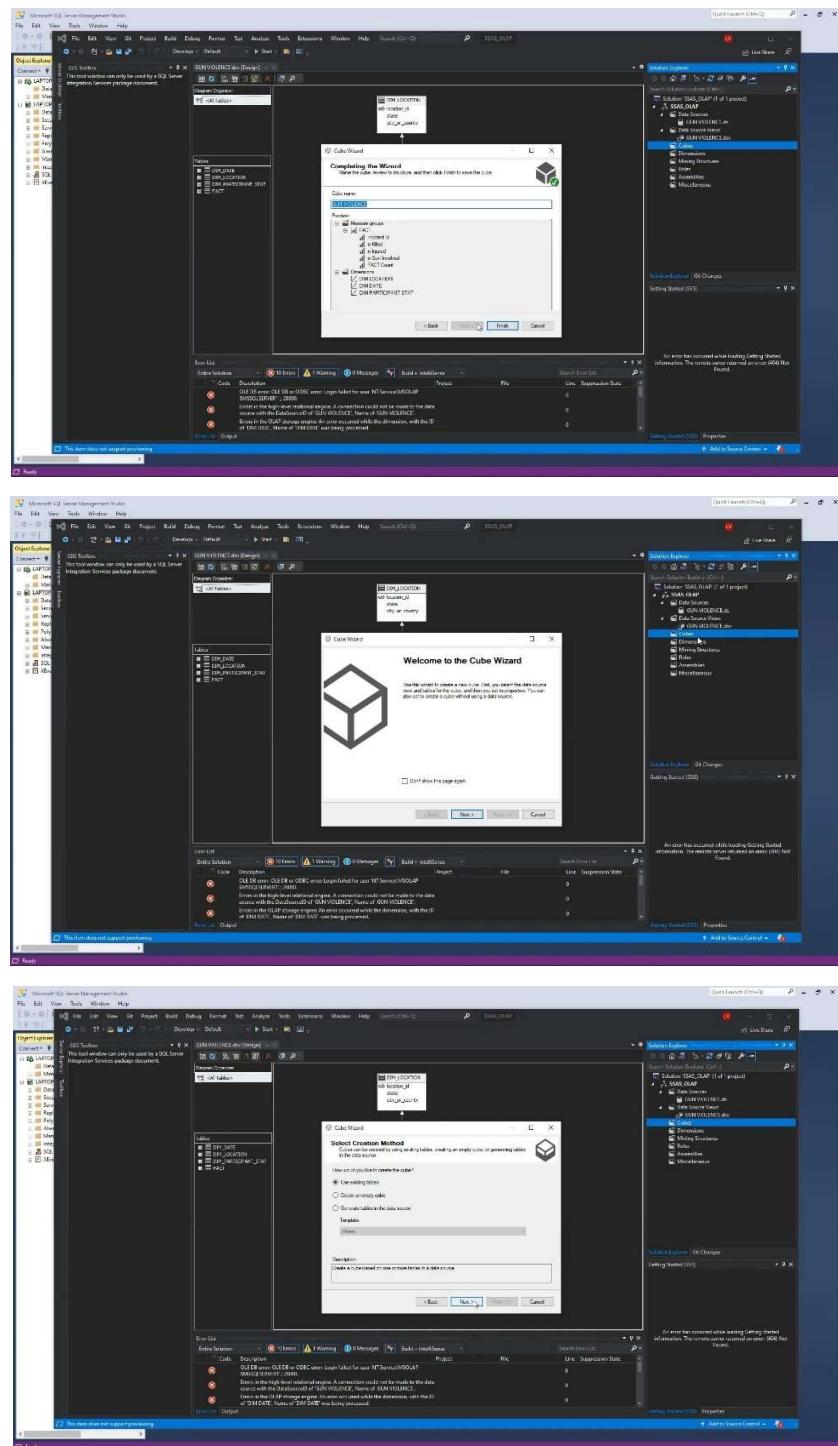


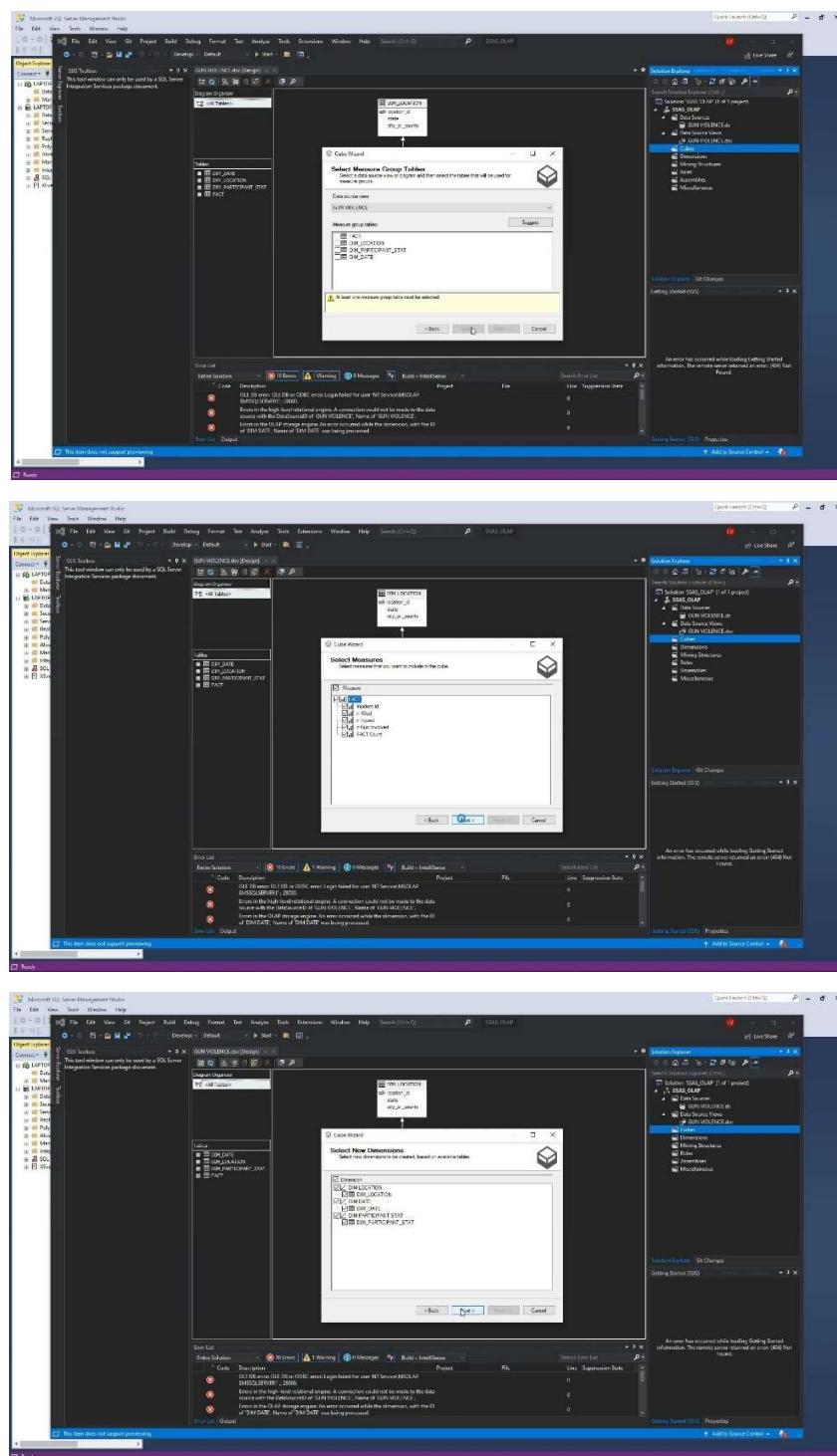
Bước 3: Tạo Data View





Bước 4: Tạo Cube và các Dimension





Bước 5: Tạo các Measures và Hierarchy (trong video Demo)

### 3. Tiến hành truy vấn 10 câu query trên Cube, MDX và Pivot Excel

Câu 1: Số vụ xả súng tại bang Nevada qua từng năm

Ý nghĩa câu truy vấn: Nevada với thành phố Las Vegas luôn là điểm nóng cho các vụ xả súng, thanh trùng lẫn nhau ở Mỹ, câu truy vấn này để thống kê xem ở Nevada có bao nhiêu vụ xả súng qua từng năm.

Kết quả truy vấn MDX:

```
SELECT{ [Measures].[FACT Count]} ON COLUMNS,
{[DIM DATE].[Year].members} ON Rows
From [GUN VIOLENCE]
WHERE [DIM LOCATION].[State_only].[State].&[Nevada];
```

FACT Count	
All	4729
2013	39
2014	3727
2015	963
Unknown	(null)

Kết quả truy vấn Cube:

Kết quả truy vấn Pivot Excel:

	FACT Count	Column Labels
Row Labels - 2013	2684 2013 Grand Total	
2013	1 209 337 479	
Grand Total	1 270 125 479	

Câu 2: Tìm ra ngày chết chóc nhất của nước Mỹ năm 2014

Ý nghĩa câu truy vấn: Tìm ra ngày có số vụ xả súng nhiều nhất ở Mỹ năm 2014 để đối chiếu với các sự kiện liên quan, từ đó rút ra kết luận xem các sự kiện đó có liên quan tới các vụ án hay không.

Kết quả truy vấn MDX:

```
select [Measures].[FACT Count] on columns,
    Head(
        Order(
            [Dim DATE].[Date].children,
            [Measures].[FACT Count], Desc) ,5) on rows
from [GUN VIOLENCE]
WHERE [DIM DATE].[Year].&[2014];
```

	FACT Count
2014-09-06	220
2014-01-01	216
2014-07-05	212
2014-10-25	210
2014-09-11	203
2014-07-13	199
2014-07-20	199
2014-08-01	199
2014-08-17	196

## Kết quả truy vấn Cube:

## Kết quả truy vấn Pivot Excel:

Câu 3: Theo từng tháng, năm liệt kê số vụ xả súng ở từng bang (Drill down)

Ý nghĩa câu truy vấn: Xem xét và so sánh đối chiếu các vụ án ở từng ban theo từng thời điểm trong năm để xem tháng nào là cao điểm của các vụ xả súng.

## Kết quả truy vấn MDX:

```
Select Non empty{[DIM LOCATION].[State].members*[Measures].[FACT Count]} on Columns,
{DrillDownLevel(
    DrillDownLevel([DIM DATE].[M_Y])
```

)  
} on Rows  
From [GUN VIOLENCE];

	All	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware	District of Columbia	Florida	Georgia	Hawaii	Idaho	Illinois	Indiana	Iowa
	FACT Count	FACT Count	FACT Count	FACT Count	FACT Count	FACT Count	FACT Count	FACT Count	FACT Count	FACT Count							
All	65535	1601	250	700	690	4729	738	744	587	959	3881	2718	66	165	3888	1526	710
2013	278	2	(null)	5	(null)	39	3	3	3	6	13	6	(null)	18	7	(null)	
1	15	(null)	(null)	(null)	2	1	(null)	(null)	1	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)
10	22	(null)	(null)	2	(null)	3	(null)	1	(null)	(null)	3	(null)	(null)	(null)	1	(null)	
11	23	(null)	(null)	1	(null)	4	(null)	(null)	(null)	1	1	1	(null)	(null)	(null)	1	(null)
12	22	1	(null)	(null)	(null)	2	(null)	(null)	1	(null)	2	(null)	(null)	(null)	(null)	(null)	(null)
2	12	(null)	(null)	(null)	(null)	4	(null)	(null)	1	(null)	(null)	1	(null)	1	(null)	1	(null)
3	19	(null)	(null)	(null)	(null)	6	(null)	(null)	(null)	1	1	1	(null)	1	1	(null)	
4	17	(null)	(null)	1	(null)	2	(null)	(null)	(null)	(null)	1	(null)	(null)	2	(null)	(null)	
5	27	(null)	(null)	1	(null)	3	(null)	(null)	(null)	(null)	(null)	1	(null)	2	1	(null)	
6	31	(null)	(null)	(null)	(null)	4	1	(null)	(null)	(null)	(null)	1	(null)	5	1	(null)	
7	34	1	(null)	(null)	(null)	4	(null)	1	(null)	1	2	(null)	(null)	3	(null)	(null)	
8	28	(null)	(null)	(null)	(null)	3	(null)	(null)	1	(null)	1	(null)	(null)	2	1	(null)	

Kết quả truy vấn Cube:

Kết quả truy vấn Pivot Excel:

Câu 4 Tên tiểu bang có số lượng người tử vong >500 trong các vụ xả súng Năm 2014

Ý nghĩa câu truy vấn: Tìm ra các điểm nóng về xả súng trong năm 2014

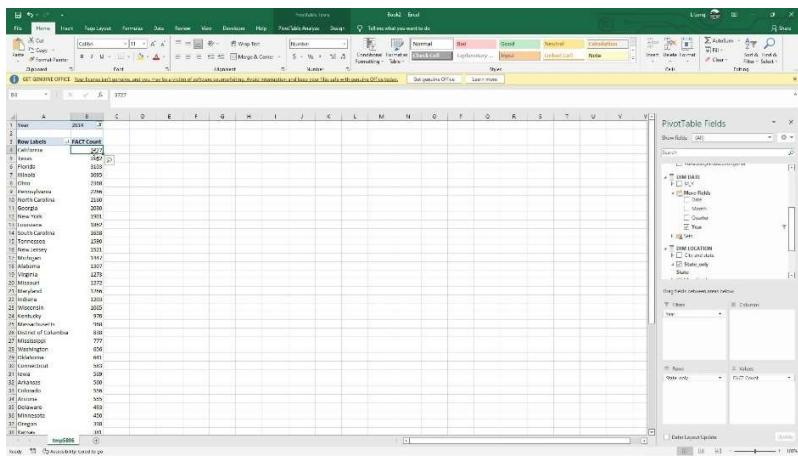
Kết quả truy vấn MDX:

```
Select {[Measures].[FACT Count]} on columns,
{FILTER ({[DIM LOCATION].[State_only].CHILDREN}, [Measures].[Soluongnguoituvong] > 500)}
on Rows
From [GUN VIOLENCE]
Where [DIM DATE].[Year].&[2014];
```

	Soluongnguoituvong
California	1320
Florida	869
Georgia	522
Illinois	637
Pennsylvania	547
Texas	1094

Kết quả truy vấn Cube:

Kết quả truy vấn Pivot Excel:



Câu 5 với mỗi tiêu ban, đưa ra 3 thành phố có số vụ xả súng nhiều nhất

Ý nghĩa câu truy vấn: Tìm ra top 3 thành phố hoặc hạt nguy hiểm nhất mỗi bang.

Kết quả truy vấn MDX:

```
Select {[Measures].[FACT count]} on Columns,
{Generate(
    [DIM LOCATION].[State_only].Children,
    TopCount({[DIM LOCATION].[State_only].CurrentMember*[DIM LOCATION].[City Or County]
    .Children},3,[Measures].[FACT Count])
)
)} on Rows
From [GUN VIOLENCE];
```

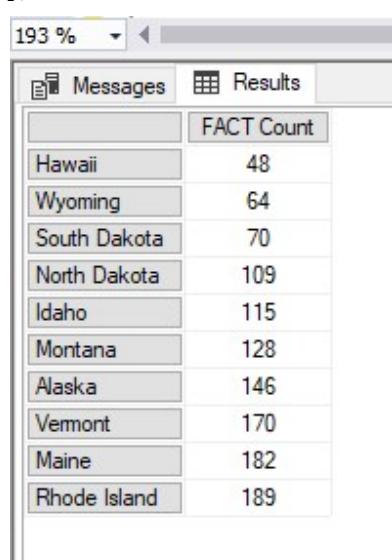
		FACT Count
Alabama	Huntsville	281
Alabama	Birmingham	210
Alabama	Mobile	153
Alaska	Anchorage	106
Alaska	Fairbanks	51
Alaska	Wasilla	16
Arizona	Phoenix	316
Arizona	Tucson	102
Arizona	Mesa	33
Arkansas	Little Rock	231
Arkansas	Jonesboro	50
Arkansas	Pine Bluff	45
California	Oakland	785
California	Los Angeles	426
California	Fresno	341

Câu 6: Tìm ra top 3 tiểu bang an toàn nhất nước Mỹ trong năm 2014

Ý nghĩa câu truy vấn: Tìm ra top các tiểu bang an toàn nhất nước Mỹ.

Kết quả truy vấn MDX:

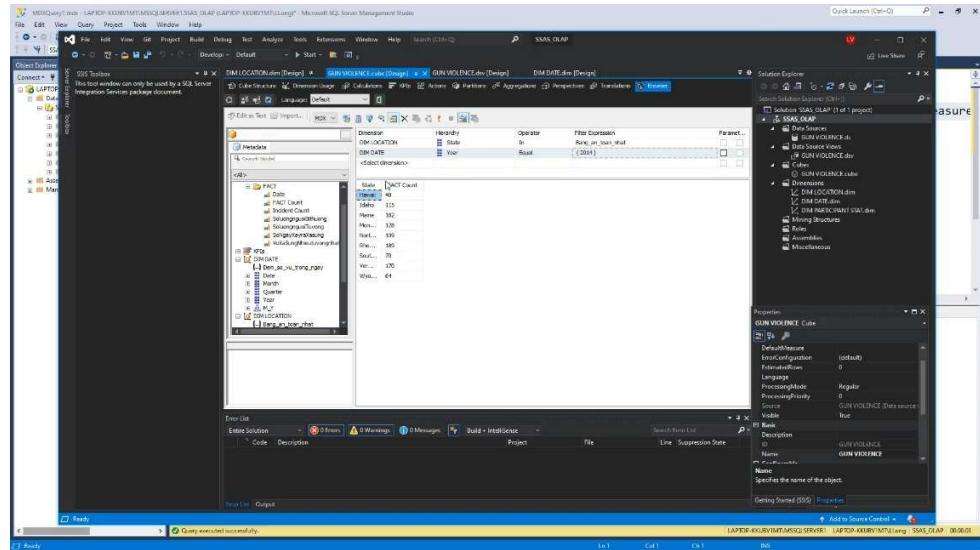
```
select {[Measures].[FACT count]} on columns,
NON EMPTY
BottomCount(NonEmpty([DIM LOCATION].[State].Children,[Measures].[FACT
count]),10,[Measures].[FACT count])
on rows
from [GUN VIOLENCE]
where [DIM DATE].[Year].&[2014];
```



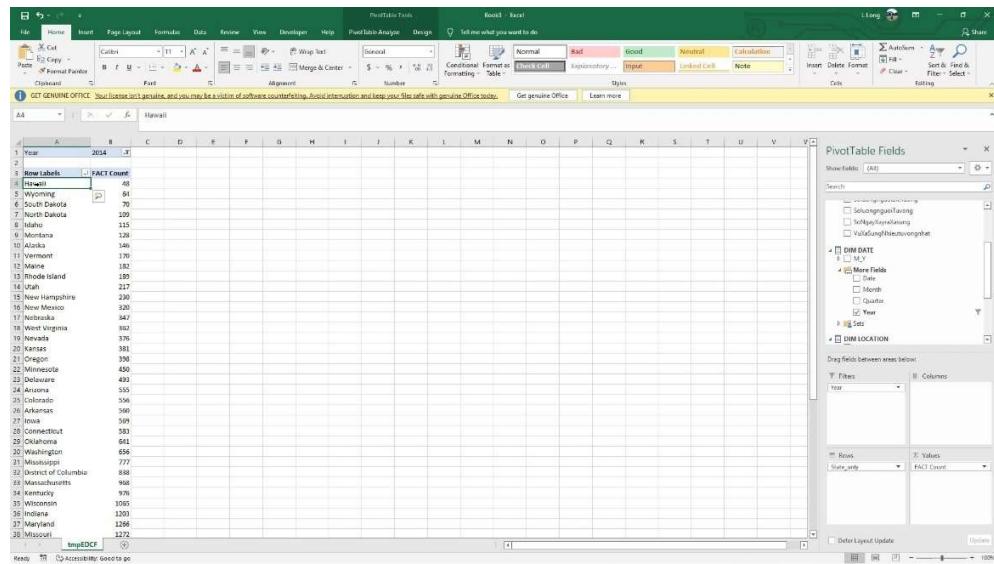
The screenshot shows the SSMS Results pane with a title bar showing '193 %'. Below the title bar, there are two tabs: 'Messages' (selected) and 'Results'. The 'Results' tab displays a table with two columns: 'State' and 'FACT Count'. The data is as follows:

	FACT Count
Hawaii	48
Wyoming	64
South Dakota	70
North Dakota	109
Idaho	115
Montana	128
Alaska	146
Vermont	170
Maine	182
Rhode Island	189

## Kết quả truy vấn Cube:



## Kết quả truy vấn Pivot Excel:



Câu 7: Tìm ra các thành phố có số vụ xả súng >10 và <20

Ý nghĩa câu truy vấn: Tìm ra các thành phố có số vụ xả súng trung bình để phục vụ cho công tác khảo sát sau này.

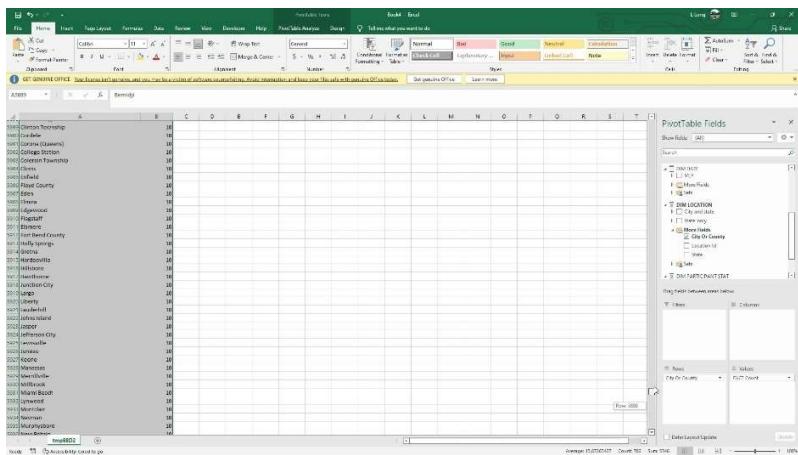
Kết quả truy vấn MDX:

```
Select {[Measures].[FACT count]} on Columns,
NON EMPTY
    {intersect(
        {FILTER([DIM LOCATION].[City Or County].CHILDREN,[Measures].[FACT
count]>=10)}
        ,{FILTER([DIM LOCATION].[City Or County].CHILDREN,[Measures].[FACT
count]<=20)}
    )} on rows
from [GUN VIOLENCE];
```

	FACT Count
Abbeville	14
Aberdeen	16
Abilene	17
Albemarle	10
Albert Lea	10
Alcoa	12
Algiers	10
Altamonte Springs	11
Alton	19
Altoona	15
Anaheim	11
Anniston	14
Apopka	18
Aransas Pass	10
Asbury Park	12

Kết quả truy vấn Cube:

Kết quả truy vấn Pivot Excel:



Câu 8: Liệt kê số vụ xả súng của các tiểu bang nằm ngoài DC

Ý nghĩa câu truy vấn: Những bát ổn về chính trị tội phạm thường nằm xa khu vực thủ đô, phép liệt kê này để so sánh số vụ án giữa các ban khác với DC.

Kết quả truy vấn MDX:

```
Select {[Measures].[FACT Count]} on columns,
{except([DIM LOCATION].[State_only].Children,[DIM LOCATION].[State_only].&[District of
Columbia])}
on Rows
```

	FACT Count
Alabama	1601
Alaska	250
Arizona	700
Arkansas	690
Californiam	4729
Colorado	738
Connecticut	744
Delaware	587
Florida	3881
Georgia	2718
Hawaii	66
Idaho	166
Illinois	3888
Indiana	1526
Iowa	710

From [GUN VIOLENCE]

Câu 9: Tính tỉ lệ % vụ xả súng có nữ liên quan.

Ý nghĩa câu truy vấn: Tính tỉ lệ % các vụ án có nữ giới liên quan (cả nạn nhân lẫn hung thủ)

Kết quả truy vấn MDX:

```
WITH
  MEMBER [Measures].[Percentage] AS
    [Measures].[Female]/
    ([Measures].[FACT Count]),
  FORMAT_STRING = 'percent'
SELECT
  NON EMPTY {[Measures].[Percentage]} ON COLUMNS,
  NON EMPTY {[DIM PARTICIPANT STAT].[GENDER]} ON ROWS
FROM [GUN VIOLENCE]
```

	Percentage
All	20.31%

Câu 10: Thống kê số vụ xả súng theo quan hệ của nạn nhân và hung thủ.

Ý nghĩa câu truy vấn: Để thống kê, so sánh xem mối liên hệ giữa nạn nhân và hung thủ nào gây ra nhiều vụ xả súng nhất.

Kết quả truy vấn MDX:

```
SELECT
  NON EMPTY {[Measures].[FACT Count]} ON COLUMNS,
  NON EMPTY {[DIM PARTICIPANT STAT].[Participant Relationship].members} ON ROWS
FROM [GUN VIOLENCE]
```

	FACT Count
All	65535
	63609
Armed Robbery	29
Co-worker	3
Drive by - Random victims	1
Family	65
Friends	22
Gang vs Gang	5
Home Invasion - Perp Knows Victim	7
Neighbor	4
Significant others - current or former	21
Aquaintance	3
Armed Robbery	4
Home Invasion - Perp Does Not Know Victim	6
Home Invasion - Perp Knows Victim	7

## Kết quả truy vấn Cube:

## Kết quả truy vấn Pivot Excel:

GENDER	COUNT
Male	4035
Female	220
Both	4255

Câu 11: Tỉ lệ giữa nam nữ trong các vụ xả súng

Ý nghĩa câu truy vấn: Để thống kê, so sánh tỉ lệ giữa giới tính của các đối tượng liên quan trong các vụ xả súng.

WITH

```

    MEMBER [Measures].[Female Percentage] AS
        [Measures].[Female]/
        ([Measures].[Female]+[Measures].[Male]),
        FORMAT_STRING = 'percent'
    MEMBER [Measures].[Male Percentage] AS
        [Measures].[Male]/
        ([Measures].[Female]+[Measures].[Male]),
        FORMAT_STRING = 'percent'
SELECT
    NON EMPTY {[Measures].[Female Percentage],[Measures].[Male Percentage]} ON COLUMNS,
    NON EMPTY {[DIM PARTICIPANT STAT].[GENDER]} ON ROWS
FROM [GUN VIOLENCE]
    
```

	Female Percentage	Male Percentage
All	11.60%	88.40%

Câu 12: Trung bình 1 ngày có bao nhiêu người chết và bị thương.

```
WITH
    MEMBER [Measures].[Killed Average] AS
        [Measures].[n Killed]/834,
        FORMAT_STRING = '###0.00'
    MEMBER [Measures].[Injured Average] AS
        [Measures].[n Injured]/834,
        FORMAT_STRING = '###0.00'
SELECT
    NON EMPTY {[Measures].[Killed Average]}, [Measures].[Injured Average]} ON COLUMNS,
    NON EMPTY {[DIM LOCATION].[State]} ON ROWS
FROM [GUN VIOLENCE]
```

	Killed Average	Injured Average
All	19.67	36.26

Câu 13. thống kê top 3 các nguyên nhân gây xả súng hàng đầu

```
select {[Measures].[Participant Relationship Count]} on columns,
NON EMPTY
    TopCount([DIM PARTICIPANT STAT].[Participant
Relationship].children,3,[Measures].[Participant Relationship Count])
on rows
from [GUN VIOLENCE]
```

	Participant Relationship Count
Family	531
Armed Robbery	453
Friends	308

Câu 14: bảng nhiều người chết vì xả súng nhất và ít nhất

```
SELECT {[Measures].[n Killed]} on columns,
UNION (
TOPCOUNT({[DIM LOCATION].[State].children},1,[Measures].[n Killed]),
BOTTOMCOUNT({[DIM LOCATION].[State].children},1,[Measures].[n Killed])) on rows
from [GUN VIOLENCE];
```

	n Killed
California	1710
Puerto Rico	(null)

Câu 15: Thống kê số vụ xả súng ở West coast

```
SELECT { [Measures].[FACT Count]} ON COLUMNS,
UNION ({[DIM LOCATION].[State].&[Alaska]}, {[DIM LOCATION].[State].&[California]}, {[DIM
LOCATION].[State].&[Hawaii]}, {[DIM LOCATION].[State].&[Oregon]}, {[DIM
LOCATION].[State].&[Washington]}) ON Rows
From [GUN VIOLENCE]
```

Messages		Results
	FACT Count	
Alaska	250	
California	4729	
Hawaii	66	
Oregon	534	
Washington	853	

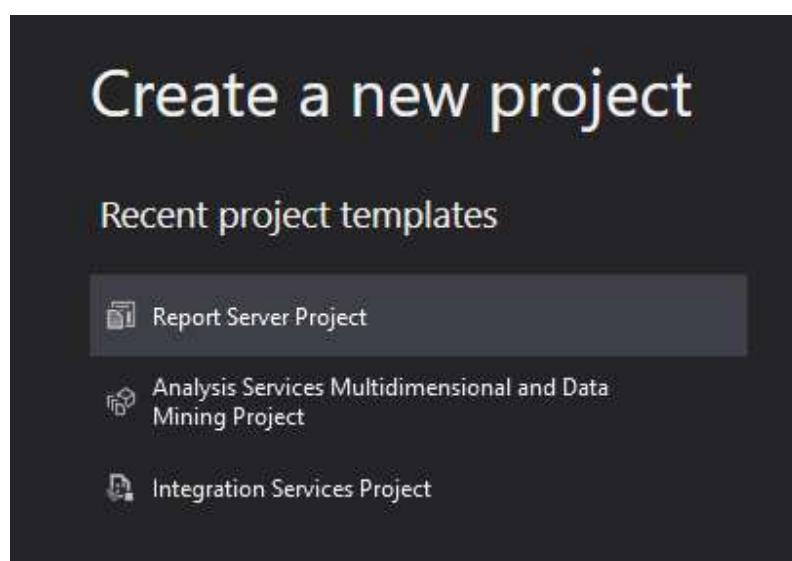
Nhận xét: Ở tất cả các câu nhóm thực hiện được truy vấn trên 3 nền tảng, kết quả truy vấn đều giống nhau.

Link sản phẩm: [https://drive.google.com/drive/folders/1XXVChnC\\_rUN9ZacinK-6w0ASK1TKH25m?usp=sharing](https://drive.google.com/drive/folders/1XXVChnC_rUN9ZacinK-6w0ASK1TKH25m?usp=sharing)

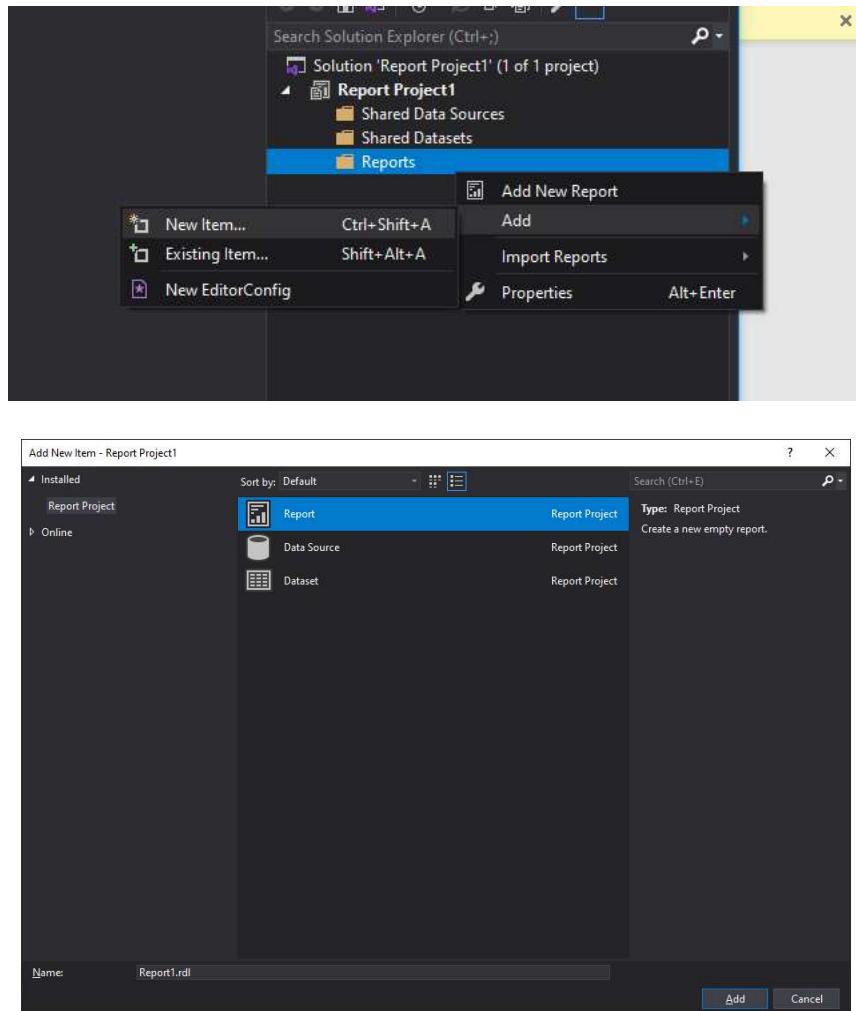
## 4. Quá trình tạo Report (SSRS)

4.1: Tạo report trên Visual Studio

Bước 1: Tạo project SSRS trong Visual Studio

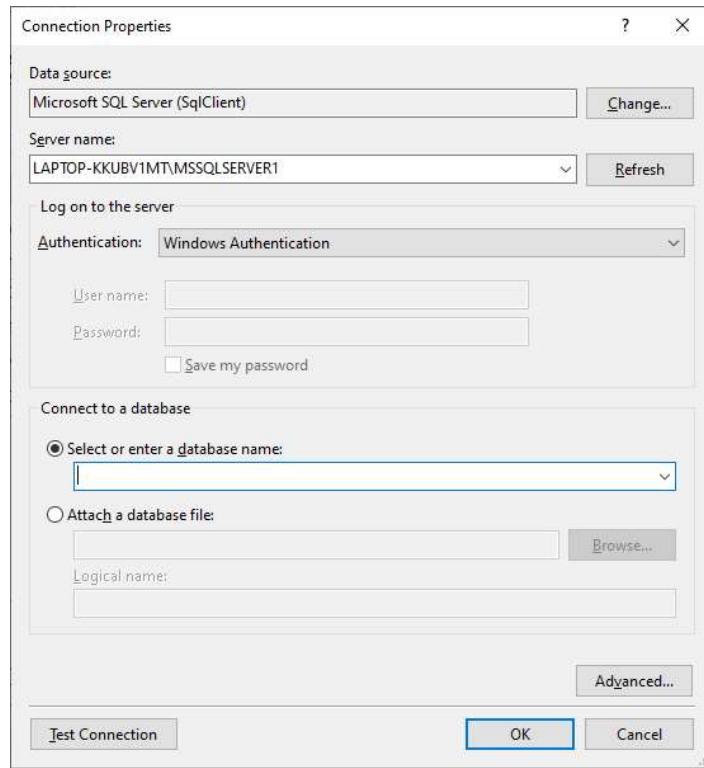


### Bước 2: Tạo report:

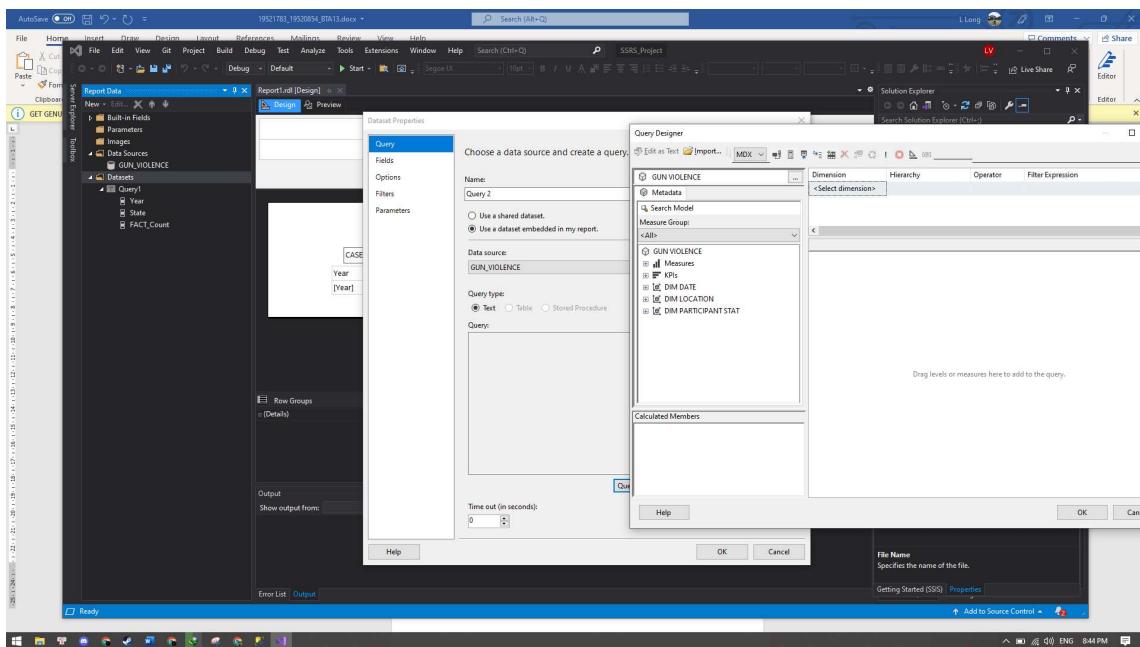


Đặt tên cho Report

### Bước 3: Tạo data source

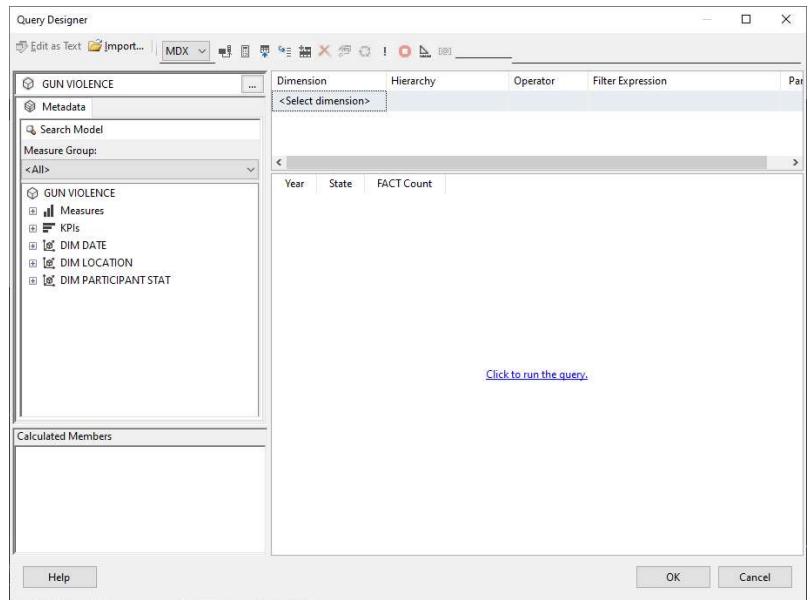


Bước 4: Tiến hành kéo thả, tạo các câu truy vấn như trong phần truy vấn trên cube trong phần dataset



### 3.2 Tiến hành tạo Report bằng project SSRS

#### 4.1.1 Report Grouping: tính tổng số vụ xả súng ở từng bang theo các năm



Query design

CASES IN EACH STATE BY YEAR

Year	State	FACT Count
2013	Alabama	2
2013	Arizona	5
2013	California	39
2013	Colorado	3
2013	Connecticut	3
2013	Delaware	3
2013	District of Columbia	6
2013	Florida	13
2013	Georgia	6
2013	Illinois	18
2013	Indiana	7
2013	Kansas	4
2013	Kentucky	3
2013	Louisiana	9
2013	Maryland	5
2013	Michigan	14
2013	Minnesota	2
2013	Mississippi	2
2013	Missouri	10

## Result

4.1.2 Report Grouping+drill down: Report về số vụ xả súng ở các bang theo từng tháng, quý năm

The screenshot shows a query design interface with a table preview and grouping definitions.

**Table Preview:**

	Year	Month	State	FACT Count
[≡]	[Year]	[Month]	[State]	[FACT_Count]

**Row Groups:**

- Year
- Month (selected)
- ≡ (Details)

**Column Groups:**

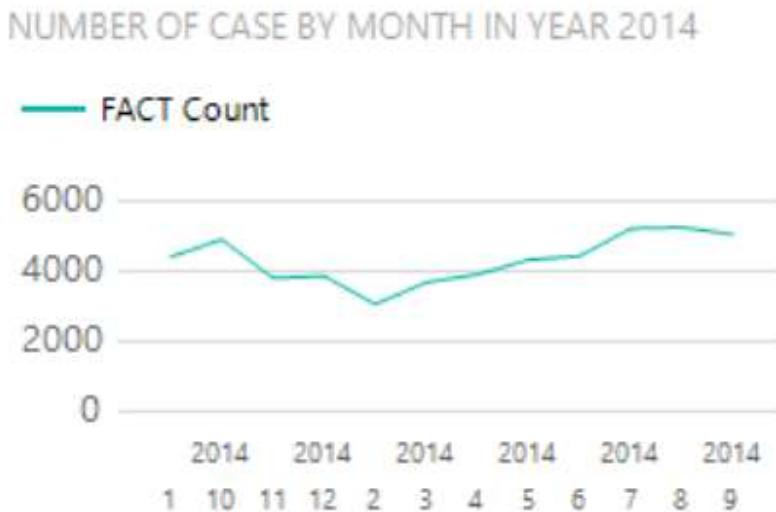
- None

Query Design

Year	Month	State	FACT Count
⇒ 2013	⇒ 1	California	2
	⇒ 10	Arizona	2
	⇒ 11	Arizona	1
	⇒ 12	Alabama	1
	⇒ 2	California	4
	⇒ 3	California	6
	⇒ 4	Arizona	1
	⇒ 5	Arizona	1
	⇒ 6	California	4
	⇒ 7	Alabama	1
	⇒ 8	California	3
	⇒ 9	California	2
⇒ 2014		Alabama	132
⇒ 2015		Alabama	86

## Result

#### 4.13 Report về độ tăng giảm của số vụ án trong từng tháng của năm 2014

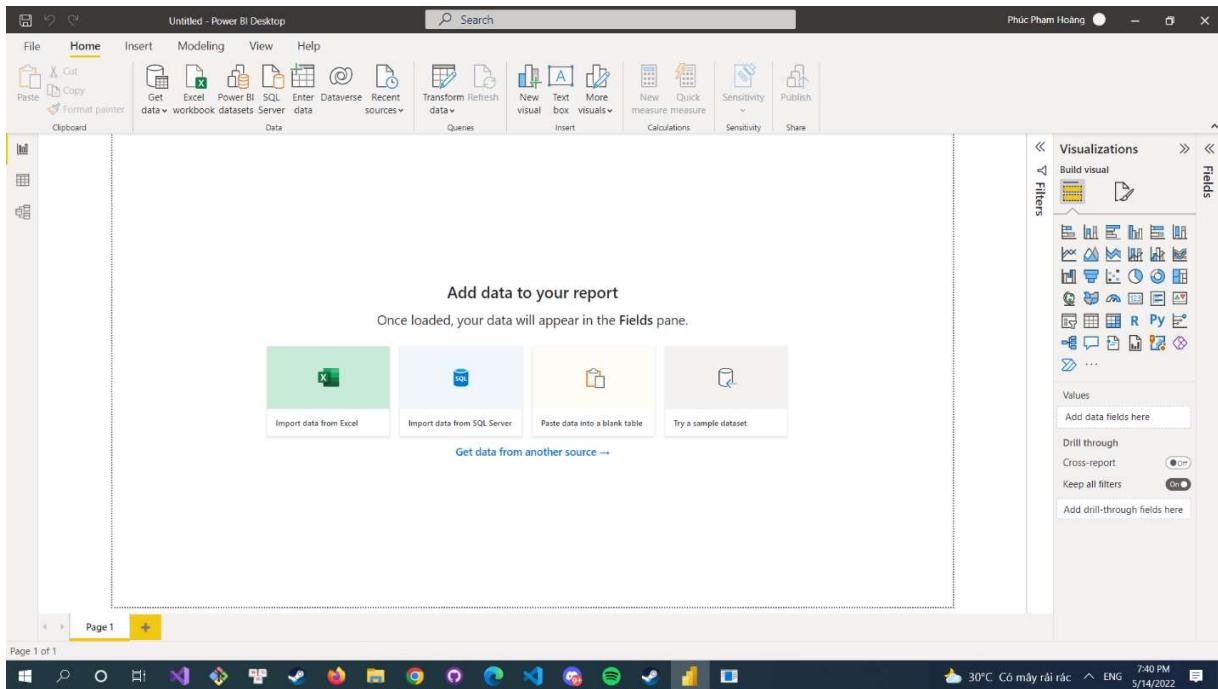


#### 4.2: Tạo Report bằng Power BI

##### 4.1.1 Cài đặt Môi trường

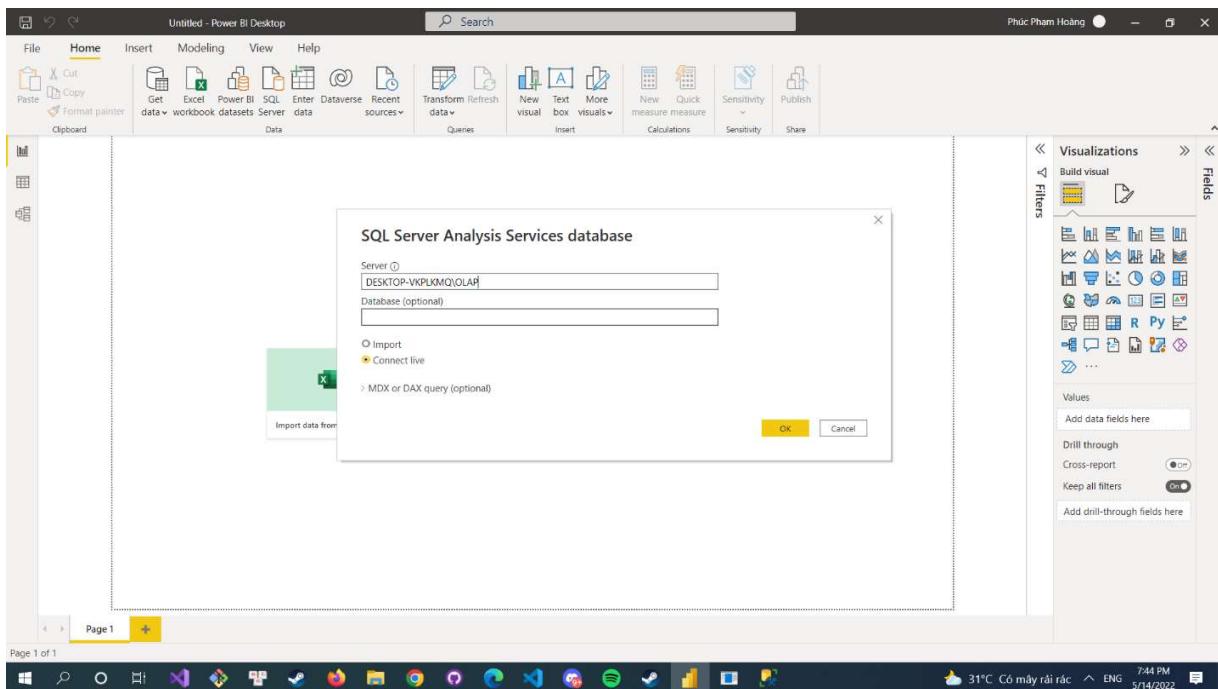
Tải PowerBI desktop từ link sau: <https://powerbi.microsoft.com/en-us/desktop/>

Ta đăng nhập tài khoản tương ứng vào PowerBI desktop sẽ được màn hình sau

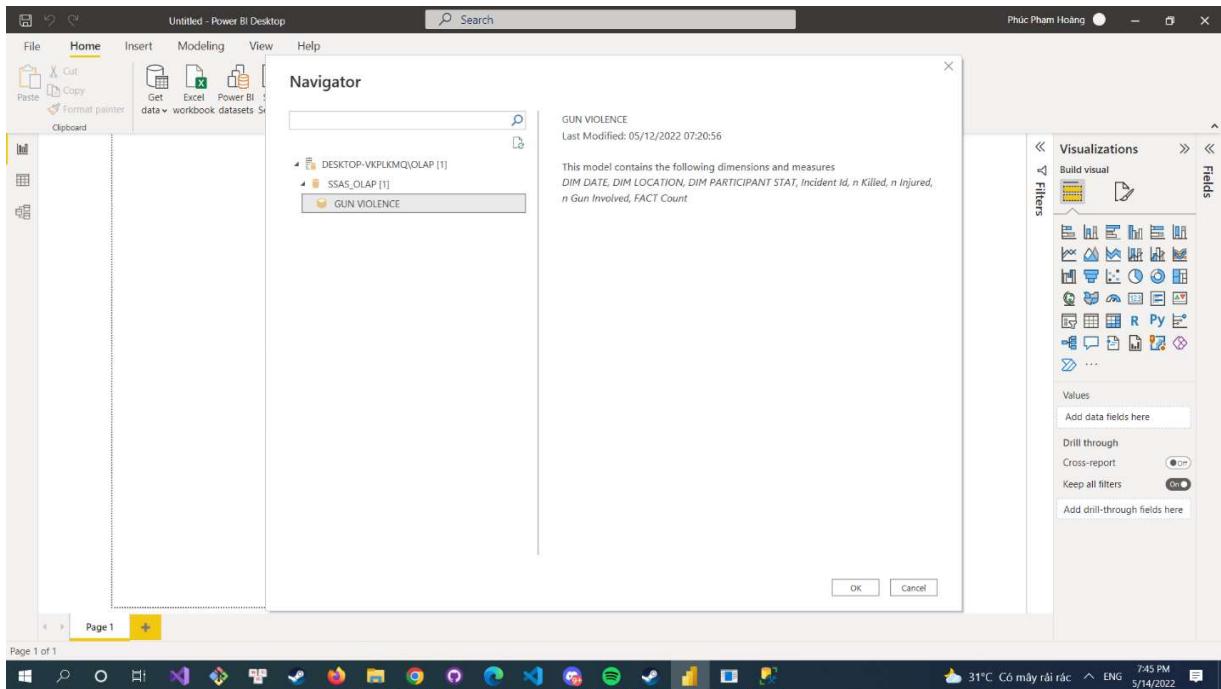


### Bước 3: Import dữ liệu từ Analysis Services

Điền tên server và chọn Connect Live



Import cube tương ứng



Các dimensions, FACT sẽ xuất hiện trong panel sau:

The screenshot shows the 'Visualizations' pane on the left and the 'Fields' pane on the right. The 'Visualizations' pane includes sections for 'Build visual' (with icons for bar, line, and pie charts), 'Values' (with a placeholder 'Add data fields here'), 'Drill through' (with options 'Cross-report' off and 'Keep all filters' on), and a placeholder 'Add drill-through fields here'. The 'Fields' pane shows a search bar and a list of dimension tables: FACT, DIM DATE, DIM LOCATION, and DIM PARTICIPANT STAT.

#### 4.1.2 Report kéo thả

Query: Số người chết vì xả súng từ 2013 - 2015

Chọn visual mong muốn



Ở đây ta chọn Line Chart

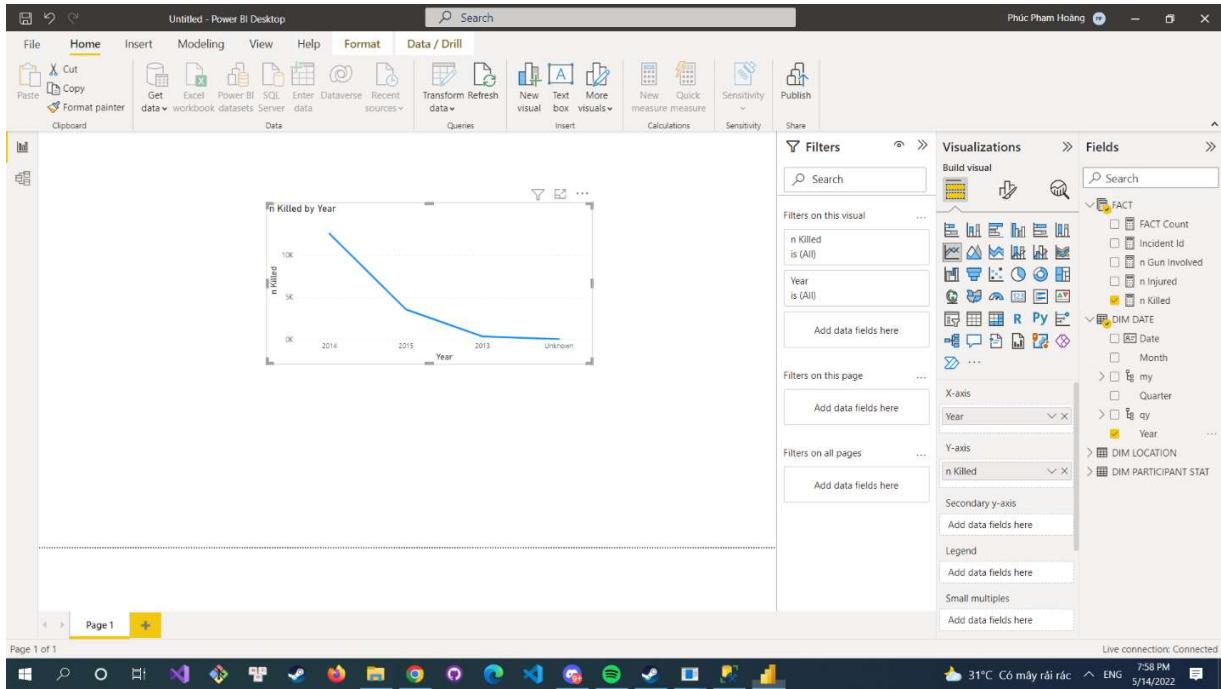
The screenshot shows the Power BI Desktop interface with the following details:

- Home Tab:** Cut, Copy, Paste, Format painter, Get data (Excel, Power BI, SQL, etc.), Transform Refresh data, New visual, Insert, More visuals, Sensitivity, Publish.
- Visualizations:** A line chart is displayed on the canvas, showing a single data series with a jagged line.
- Fields:** The Fields pane on the right lists several data tables and their fields:
  - FACT:** FACT Count, a Gun involved, a Injured, a Killed.
  - DIM DATE:** Date, Month, Year.
  - DIM LOCATION:**
  - DIM PARTICIPANT STAT:**
- Page:** Page 1 of 1.
- System Status:** Live connection Connected, 31°C, Có mây rì rác, ENG, 7:57 PM, 5/14/2022.

Kéo Hierarchy qy vào x-axis và n-killed vào y-axis



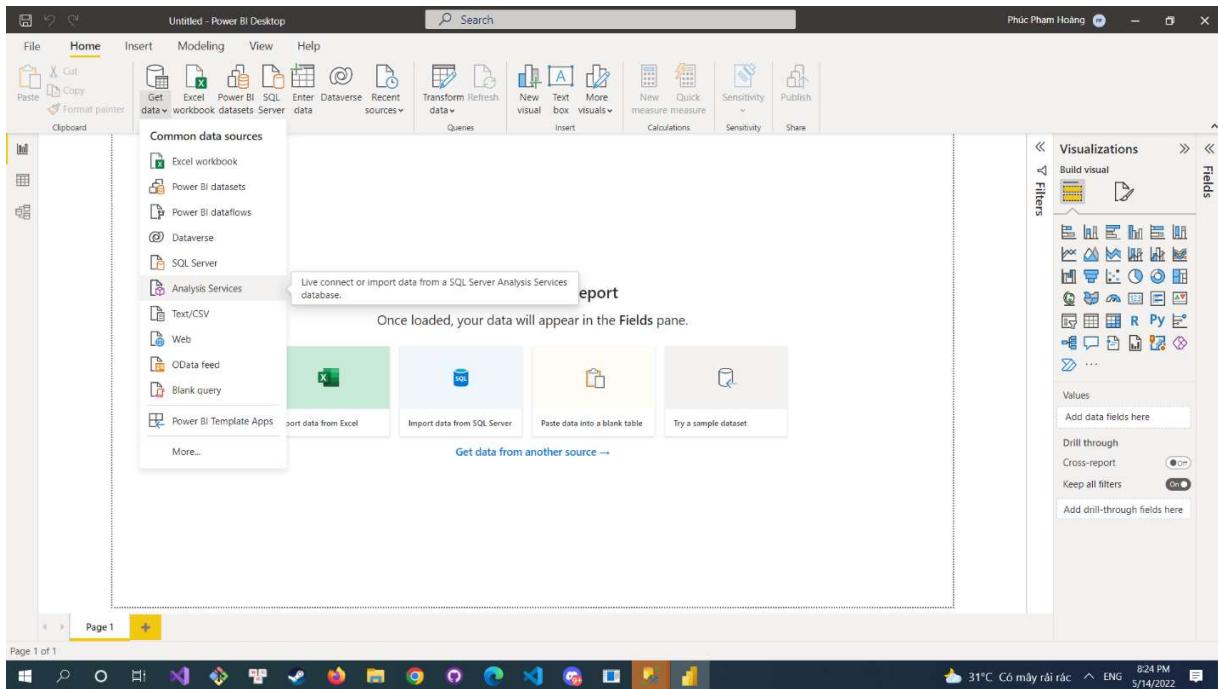
Ta sẽ được line chart như sau



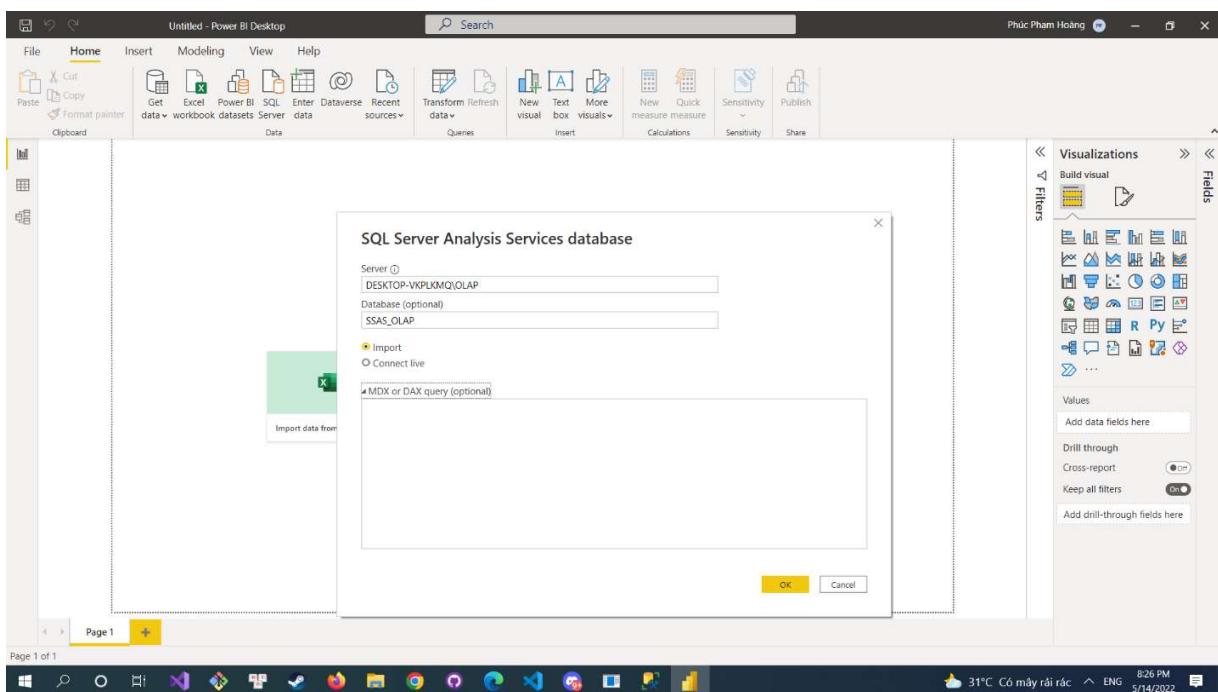
#### 4.1.3 Report bằng MDX

Câu query: Liệt kê số vụ xả súng của các tiểu bang ngoài bang DC

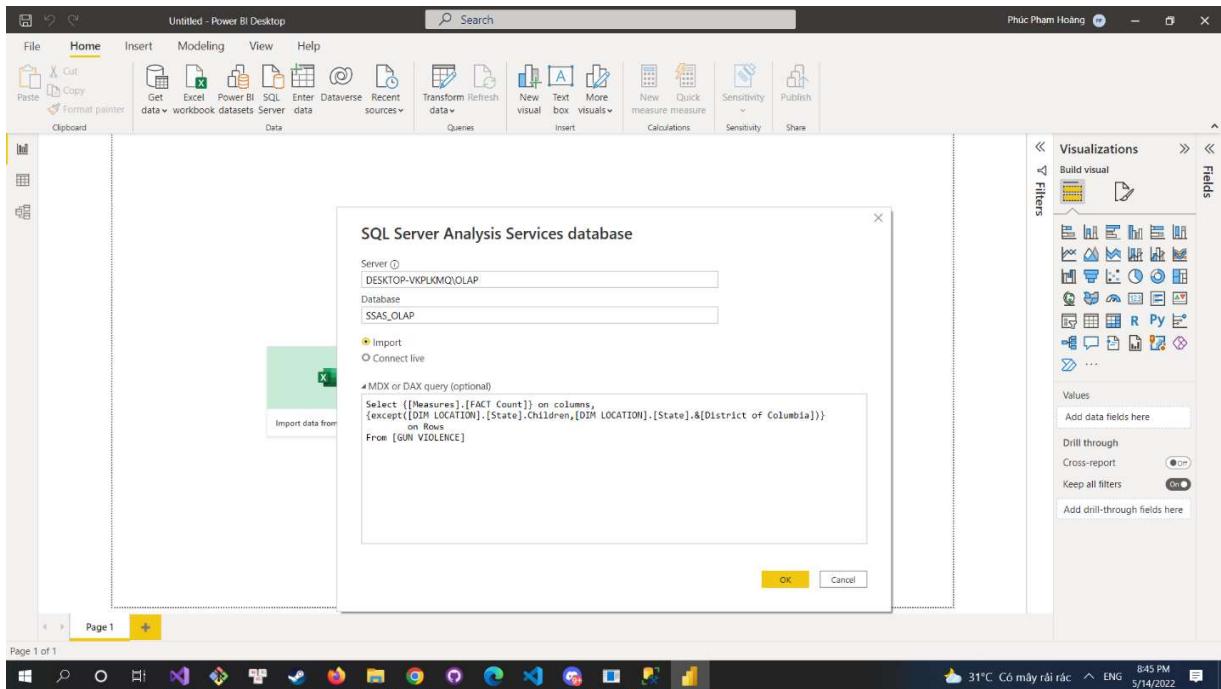
Chọn Get Data – Chọn Analysis Services



Điền tên server và điền tên database tương ứng



Điền MDX Query



Ta sẽ có data như sau:

## DESKTOP-VKPLKMQ\OLAP: SSAS\_O LAP

[DIM LOCATION].[State].[State].[MEMBER_CAPTION]	[Measures].[FACT Count]
Alabama	1601
Alaska	250
Arizona	700
Arkansas	690
California	4729
Colorado	738
Connecticut	744
Delaware	587
Florida	3881
Georgia	2718
Hawaii	66
Idaho	166
Illinois	3888
Indiana	1526
Iowa	710
Kansas	499
Kentucky	1203
Louisiana	2299
Maine	240
Maryland	1519

i The data in the preview has been truncated due to size limits.

Load Transform Data

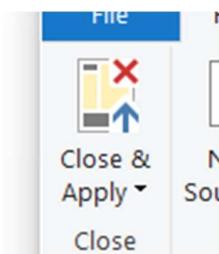
Hiện data có tên column khá xấu, ta sẽ modify lại bằng cách trọn transform data

	A <sup>B</sup> <sub>C</sub> State	ABC Number of Gun Violences
1	Alabama	1601
2	Alaska	250
3	Arizona	700
4	Arkansas	690
5	California	4729
6	Colorado	738
7	Connecticut	744
8	Delaware	587
9	Florida	3881
10	Georgia	2718
11	Hawaii	66
12	Idaho	166
13	Illinois	3888
14	Indiana	1526
15	Iowa	710
16	Kansas	499
17	Kentucky	1203
18	Louisiana	2299
19	Maine	240
20	Maryland	1519

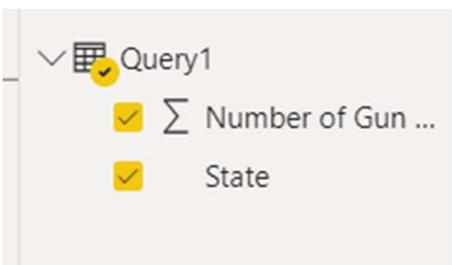
Ta cũng transform type cho các measure sang các type phù hợp, chẳng hạn như Count ở đây sẽ sang Decimal

The screenshot shows the Power Query Editor interface. A context menu is open over the 'Number of Gun Violences' column in the 'Query1' table. The 'Change Type' option is selected, with a submenu showing various numeric types: Decimal Number, Fixed decimal number, Whole Number, Percentage, Date/Time, Date, Time, Duration, Text, True/False, Binary, and Using Locale... The 'Applied Steps' pane on the right shows the 'Renamed Columns' step applied to 'Query1'. The status bar at the bottom indicates 'PREVIEW DOWNLOADED AT 8:45 PM'.

Chọn Close & Apply để hoàn tất đồ dữ liệu từ DMX query



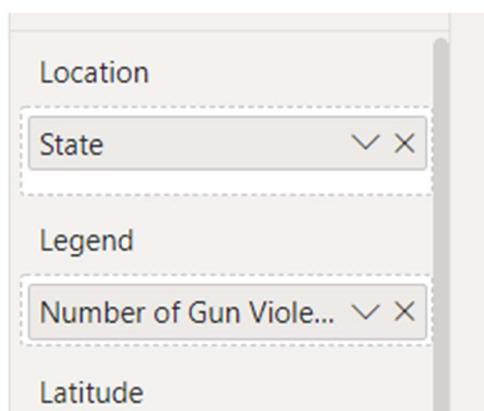
Dữ liệu đã được đồ hoàn tất, ta có thể kiểm tra bên panel fields



Do data ta là state của nước Mỹ, ta có thể tận dụng Filled Map của PowerBI (tận dụng tốt với nước Mỹ)

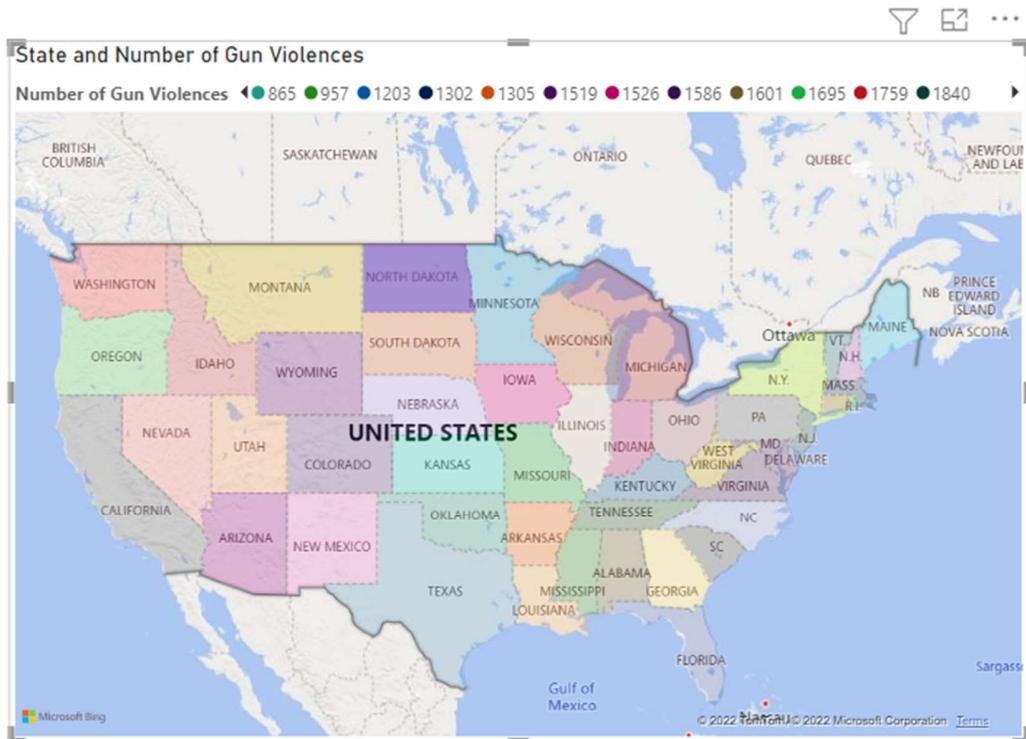


Kéo thả các thành phần trong fields vô chổ phù hợp

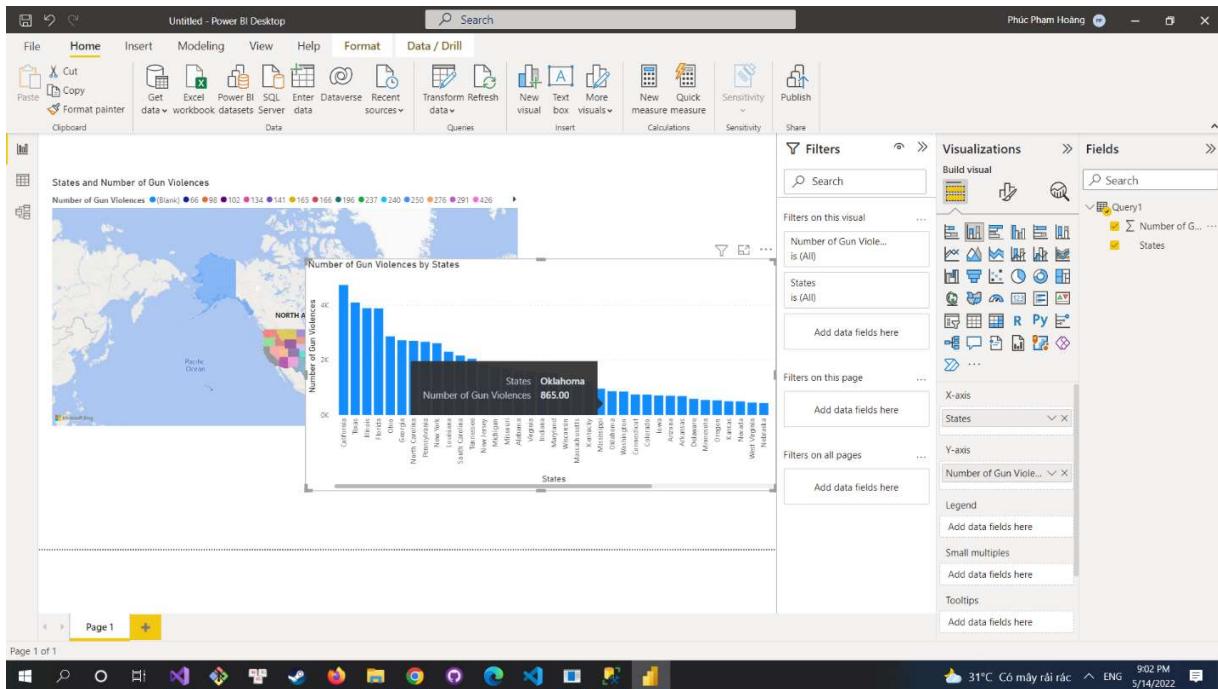


Ta sẽ được bản đồ nhiệt phía dưới với các ký tự đánh số với mức độ màu khác nhau

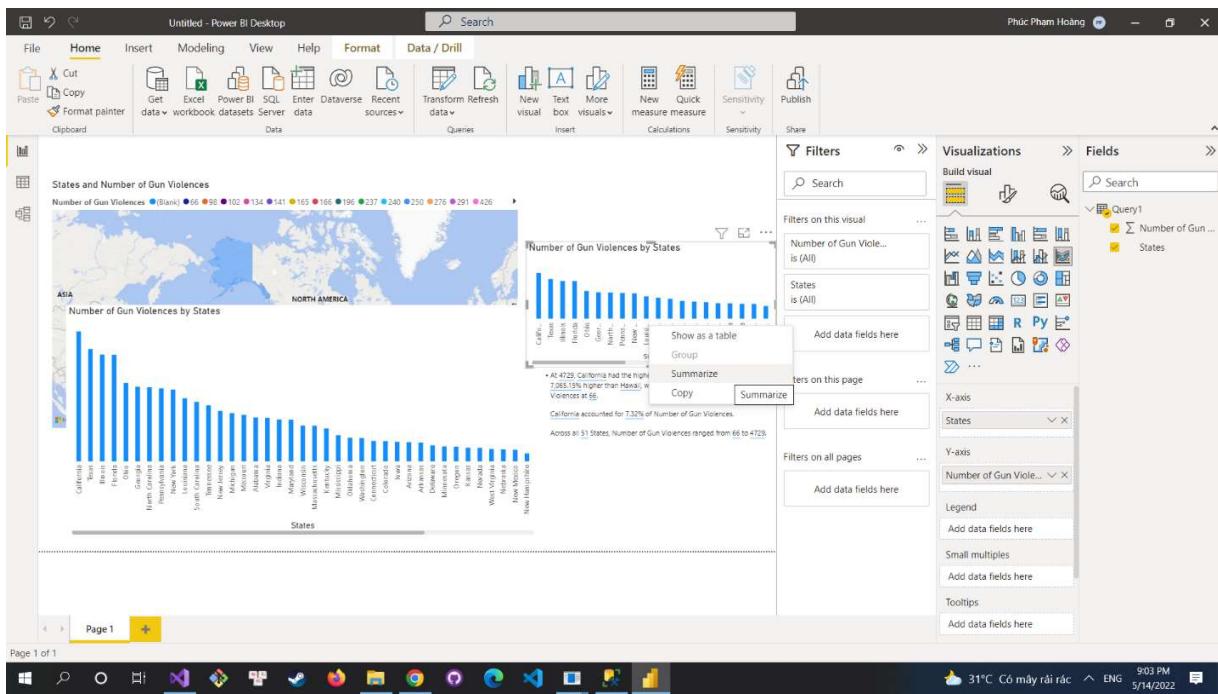
,lnm\XZ9VI



### Stacked Column Chart



Ta còn có thể chuột phải, chọn summarize và sẽ được generate sẵn một ô text có thể chỉnh sửa về nội dung của chart đó



#### 4.1.4 Pie Chart

Câu query: Câu 10 Thống kê số vụ xả súng theo quan hệ của nạn nhân và hung thủ

▪ MDX or DAX query (optional)

```
SELECT
NON EMPTY {[Measures].[FACT Count]} ON COLUMNS,
NON EMPTY {[DIM PARTICIPANT STAT].[Participant Relationship].members} ON ROWS
FROM [GUN VIOLENCE]
```

## Transform data

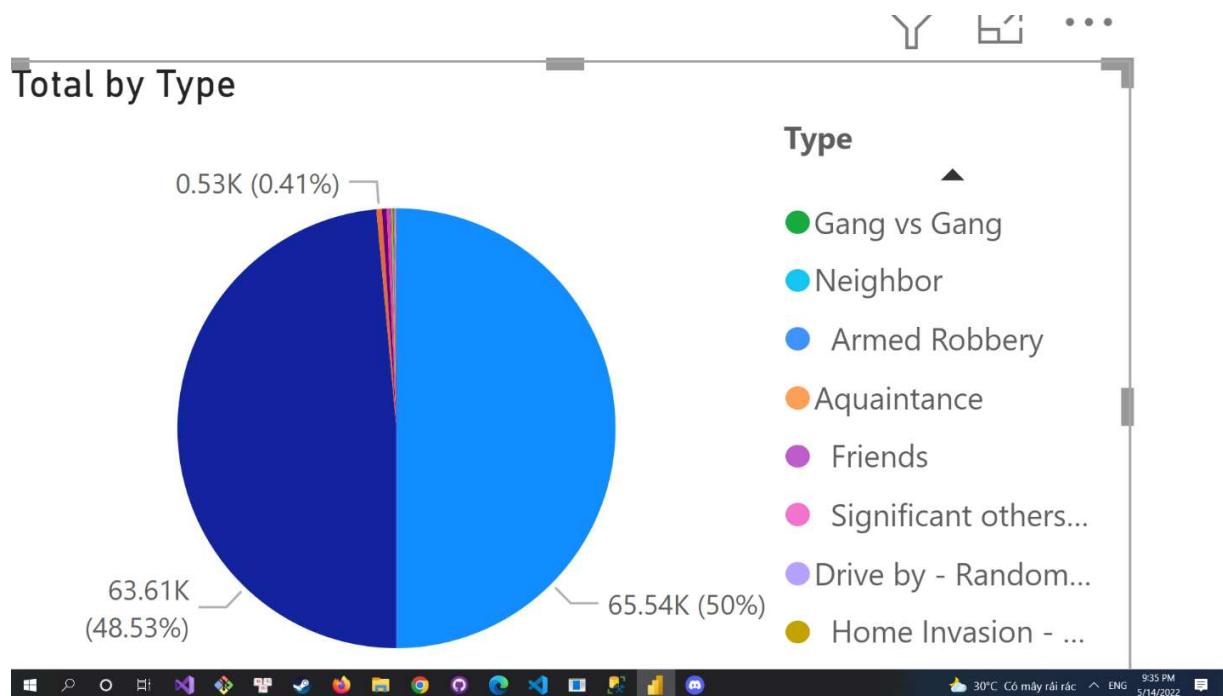
The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table titled "Type" with 20 rows and 2 columns. The columns are labeled "Type" and "Total". The data includes various types of incidents such as Armed Robbery, Co-worker, Family, Friends, Gang vs Gang, Home Invasion - Perp Knows Victim, Neighbor, Significant others - current or former, Acquaintance, Armed Robbery, Home Invasion - Perp Does Not Know Victim, Home Invasion - Perp Knows Victim, and Armed Robbery4. The total count for all rows is 65335.

The "APPLIED STEPS" pane on the right shows the history of transformations:

- Query2
- Renamed Columns
- Changed Type
- Replaced Value

The "PROPERTIES" pane shows the query is named "Query2".

result



#### 4.1.5 Grouping and Total 1

Query: với mỗi tiểu bang, đưa ra 3 thành phố có số vụ xả súng nhiều nhất

Import data

SQL Server Analysis Services database

Server

Database

Import  Connect live

MDX or DAX query (optional)

```
Select {[Measures].[FACT count]} on Columns,
{Generate(
    [DIM LOCATION].[State].Children,
    TopCount({
        [DIM LOCATION].[State].CurrentMember*[DIM LOCATION].[City Or County] .Children
        ,3,[Measures].[FACT Count]
    })
)} on Rows
From [GUN VIOLENCE];
S
```

**OK** **Cancel**

Untitled - Power BI Desktop

File Home Insert Modeling View Help

Clipboard Get data from Power BI

DESKTOP-VKPLKMQ\OLAP: SSAS\_O LAP

[DIM LOCATION].[State].[State].[MEMBER_CAPTION]	[DIM LOCATION].[City Or County].[City Or County].[MEASURES].[FACT Count]
Alabama	Huntsville
Alabama	Birmingham
Alabama	Mobile
Alaska	Anchorage
Alaska	Fairbanks
Alaska	Wasilla
Arizona	Phoenix
Arizona	Tucson
Arizona	Mesa
Arkansas	Little Rock
Arkansas	Jonesboro
Arkansas	Pine Bluff
California	Oakland
California	Los Angeles
California	Fresno
Colorado	Denver
Colorado	Colorado Springs
Colorado	Pueblo
Connecticut	New Haven
Connecticut	Hartford

The data in the preview has been truncated due to size limits.

Load Transform Data Cancel

Visualizations Fields

Query1 Number of Gun ... States

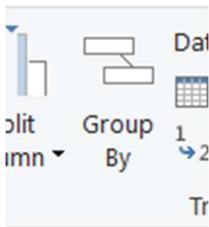
Query2 Count Relationship Type

data fields here through report all filters drill-through fields here

Page 3 of 3

30°C Có mây rải rác 10:10 PM ENG 5/14/2022

Transform Data – Chọn group by



Chọn group by theo state

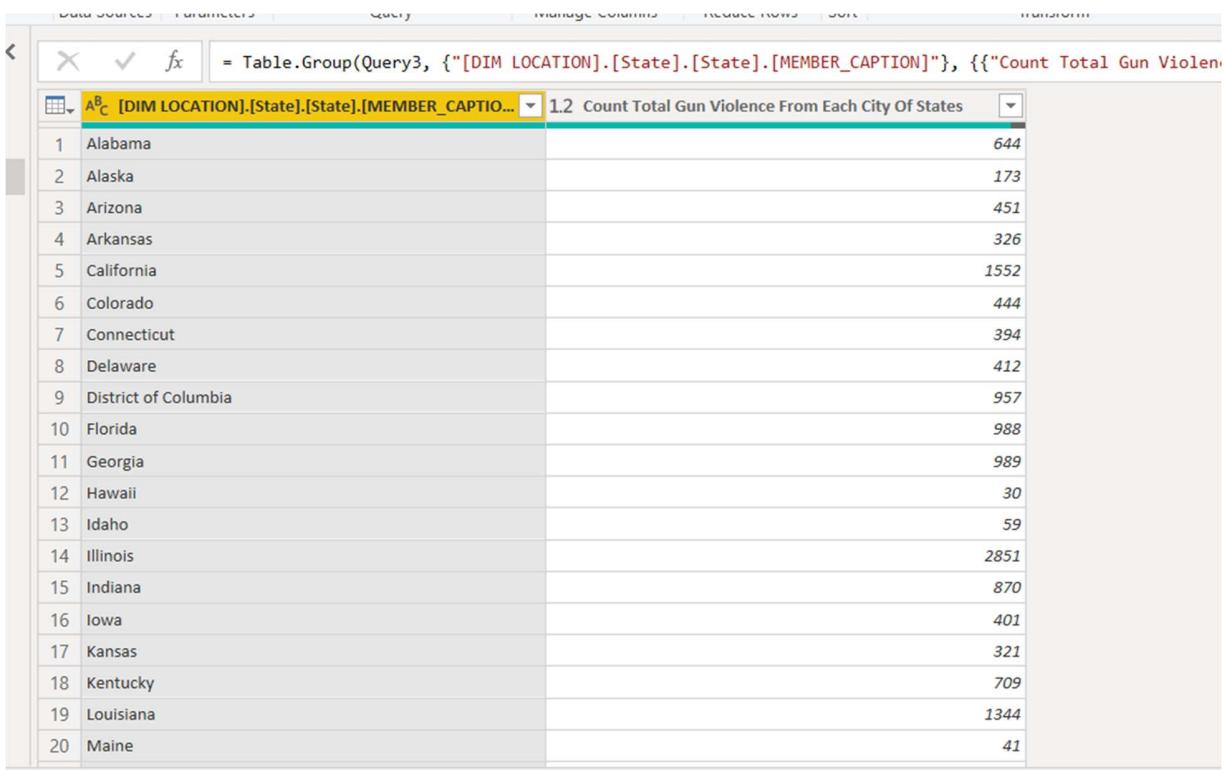
## Group By

Specify the column to group by and the desired output.

Basic     Advanced

New column name	Operation	Column
Count Total Gun Violence From Each City Of States	Sum	[Measures].[FACT Count]

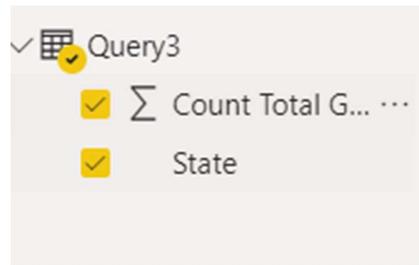
Lúc này sẽ sum (Total) lại tất cả fact count tương ứng của từng thành phố thuộc state đó và lưu vào cột mới có tên là Count Total Gun Violence From Each City Of States



A screenshot of the Power BI Data View window. The table displays the total count of gun violence incidents for each US state. The columns are labeled 'State' and 'Count Total Gun Violence'. The data shows that Illinois has the highest count at 2851, followed by Florida at 988, and Georgia at 989. The full table is as follows:

	Count Total Gun Violence
1 Alabama	644
2 Alaska	173
3 Arizona	451
4 Arkansas	326
5 California	1552
6 Colorado	444
7 Connecticut	394
8 Delaware	412
9 District of Columbia	957
10 Florida	988
11 Georgia	989
12 Hawaii	30
13 Idaho	59
14 Illinois	2851
15 Indiana	870
16 Iowa	401
17 Kansas	321
18 Kentucky	709
19 Louisiana	1344
20 Maine	41

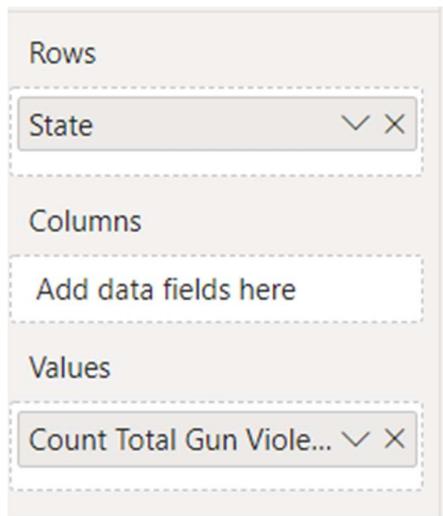
Close & Apply ta sẽ có các mục trong Fields:



Chọn Matrix



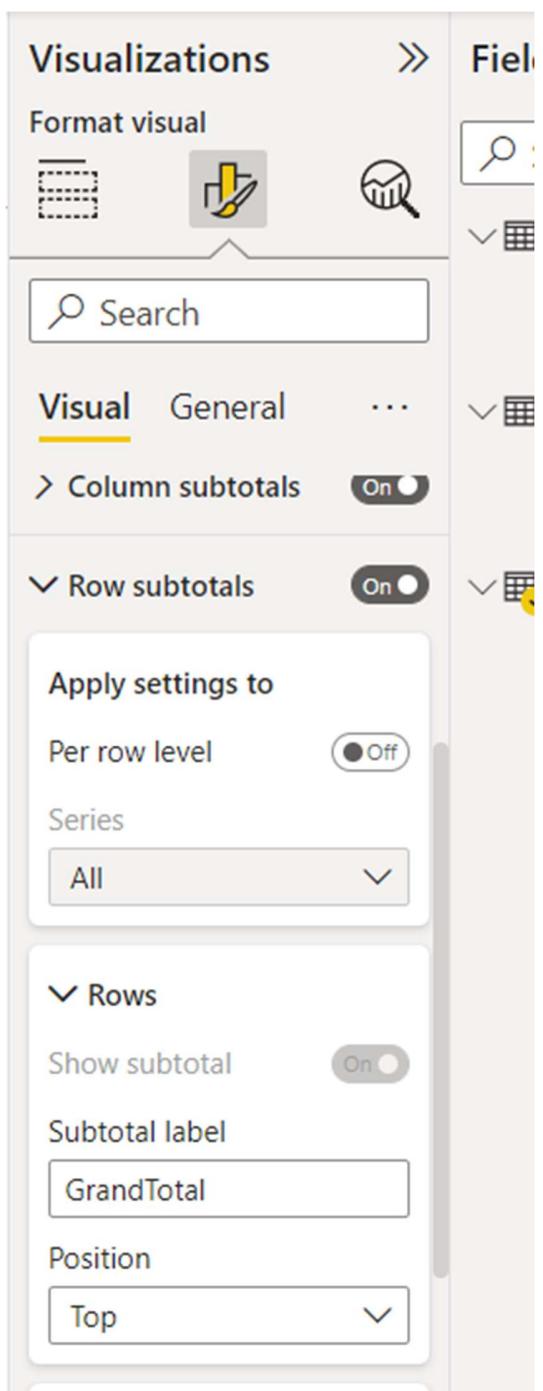
Xếp các field vào Rows và Values tương ứng



Ta sẽ được bảng có cả dòng Total(GrandTotal)

State	Count Total Gun Violence From Each City Of States
Alabama	644.00
Alaska	173.00
Arizona	451.00
Arkansas	326.00
California	1,552.00
Colorado	444.00
Connecticut	394.00
Delaware	412.00
District of Columbia	957.00
Florida	988.00
Georgia	989.00
<b>Total</b>	<b>31,047.00</b>

Ta cũng có thể rename cả label của Total bằng cách vô chỉnh phần sau



#### 4.1.6 Grouping and Total 2

Query: Số vụ xả súng xảy ra ở từng bang mỗi năm

#### SQL Server Analysis Services database

Server ⓘ

DESKTOP-VKPLKMQ\OLAP

Database

SSAS\_OLAP

Import

Connect live

▲ MDX or DAX query (optional)

```
Select {[Measures].[FACT count]} on Columns,
{Generate(
    [DIM DATE].[Year].Children,
    TopCount({
        [DIM DATE].[Year].CurrentMember*[DIM LOCATION].[State].Children}
       ,53,[Measures].[FACT Count]
    )
)} on Rows
From [GUN VIOLENCE];
```

## DESKTOP-VKPLKMQ\OLAP: SSAS\_O LAP

[DIM DATE].[Year].[Year].[MEMBER_CAPTION]	[DIM LOCATION].[State].[State].[MEMBER_CAPTION]	[Measures].[FACT Count]
2013	California	39
2013	Pennsylvania	20
2013	Illinois	18
2013	New York	16
2013	Michigan	14
2013	New Jersey	14
2013	Florida	13
2013	North Carolina	12
2013	Texas	11
2013	Missouri	10
2013	Ohio	10
2013	Tennessee	10
2013	Louisiana	9
2013	Indiana	7
2013	District of Columbia	6
2013	Georgia	6
2013	Oklahoma	6
2013	Virginia	6
2013	Arizona	5
2013	Maryland	5

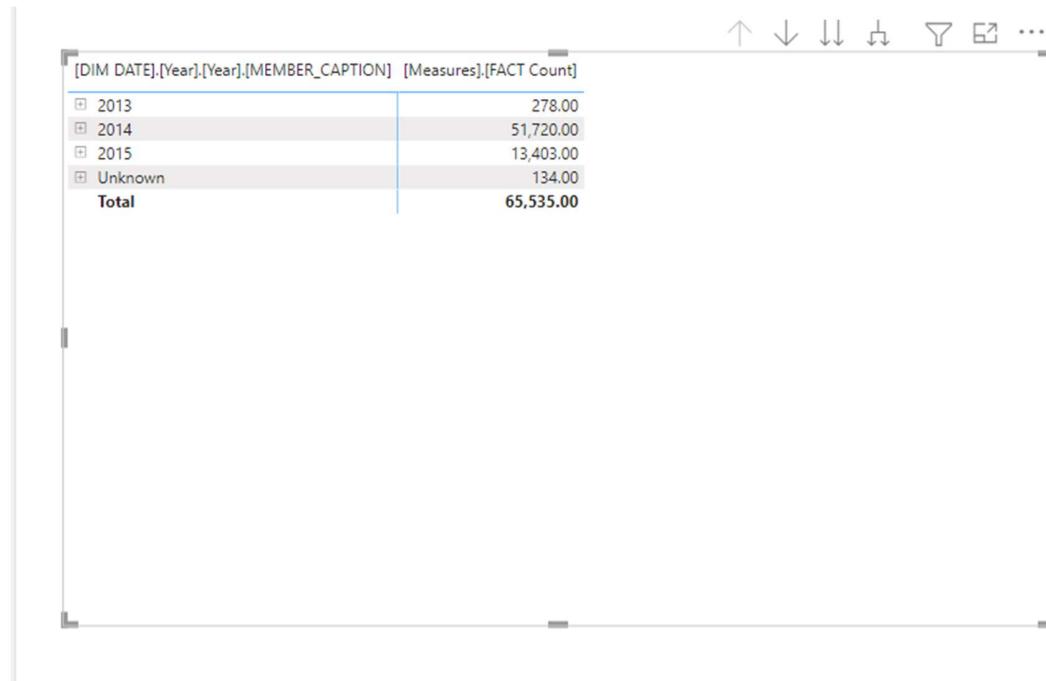
The data in the preview has been truncated due to size limits.

Load Transform Data Cancel

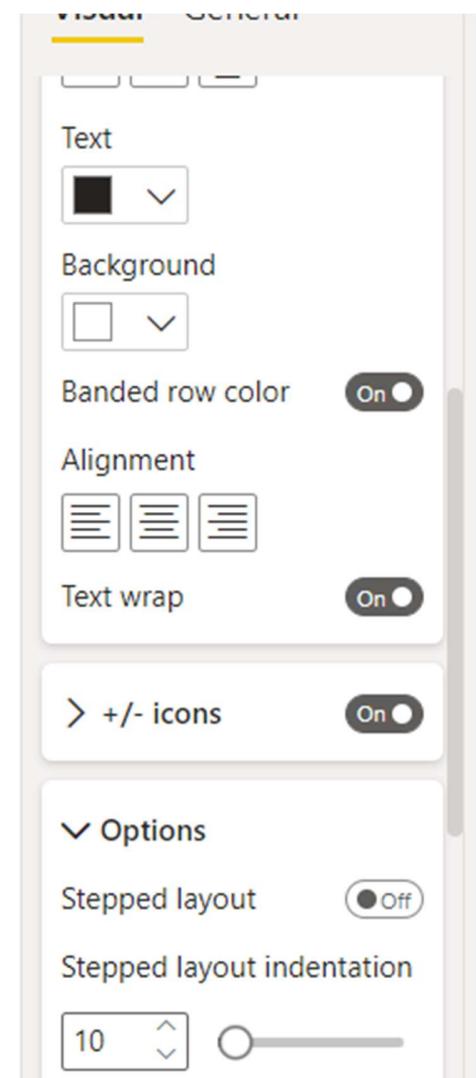
Data này ta chỉ cần tiến hành transform count sang decimal grouping Matrix visualization sẽ giúp ta

1.2 [Measures].[FACT Count]
39
20
18
16
14
14
13
12
11
10
10
10
9
7
6
6
6
5
5

### Data ta có dạng



Nhiệm vụ ta là bỏ vào đúng Rows, Values và tích Stepped Layout = Off



Click vào dấu + nhỏ kẽ năm ta sẽ được data group theo đúng ý

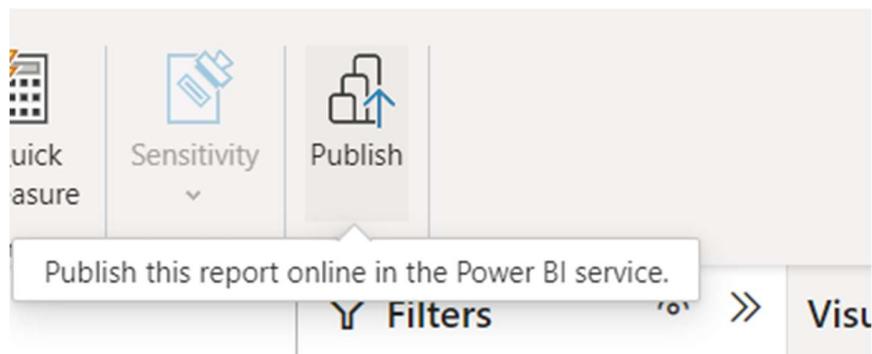
[DIM DATE].[Year].[Year].[MEMBER_CAPTION] [DIM LOCATION].[State].[State].[MEMBER_CAPTION] [Measures].[FACT Count]		
2013	Alabama	2.00
	Alaska	
	Arizona	5.00
	Arkansas	
	California	39.00
	Colorado	3.00
	Connecticut	3.00
	Delaware	3.00
	District of Columbia	6.00
	Florida	13.00
	Georgia	6.00
	Hawaii	
	Idaho	
	Illinois	18.00
	Indiana	7.00
	Iowa	
	Kansas	4.00
	Kentucky	3.00
	Louisiana	9.00
	Maine	
	Maryland	5.00

Total cũng đã được tính

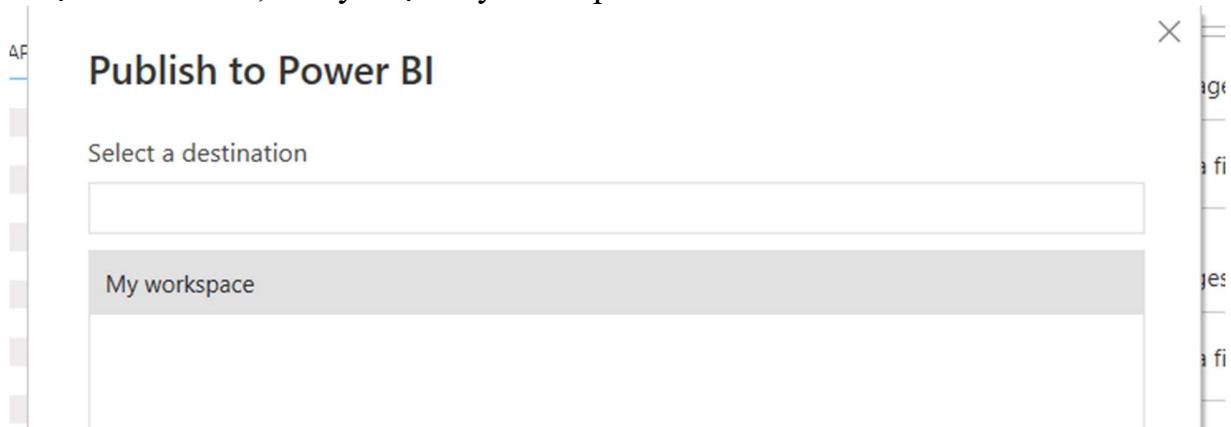
Total	65,535.00
-------	-----------

#### 4.1.7 Publish Report lên PowerBI Services

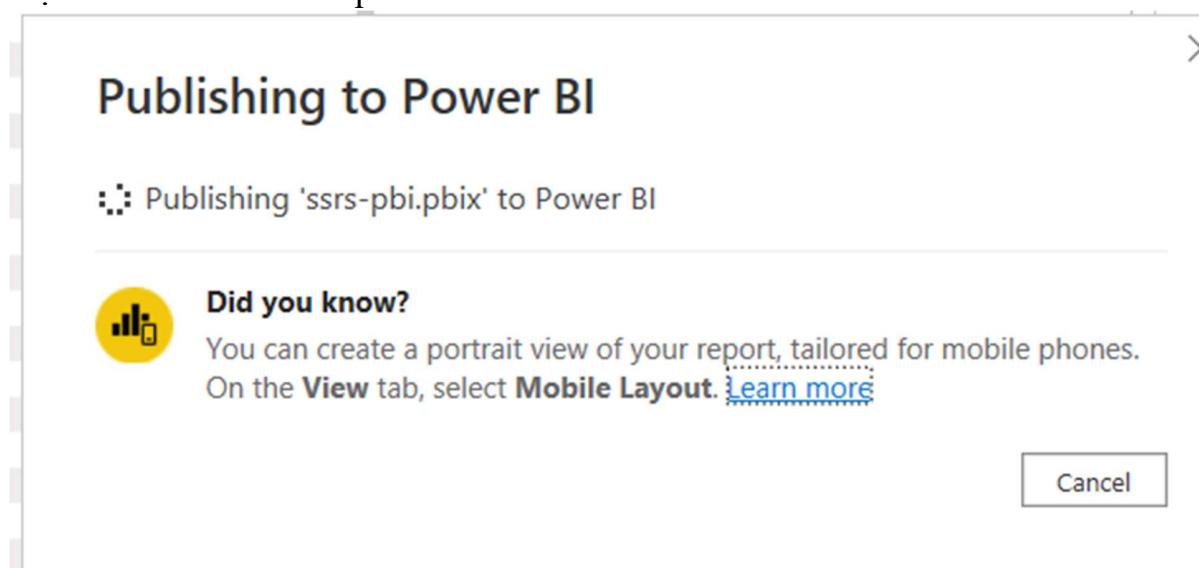
Chọn publish



Chọn destination, ở đây chọn My workspace



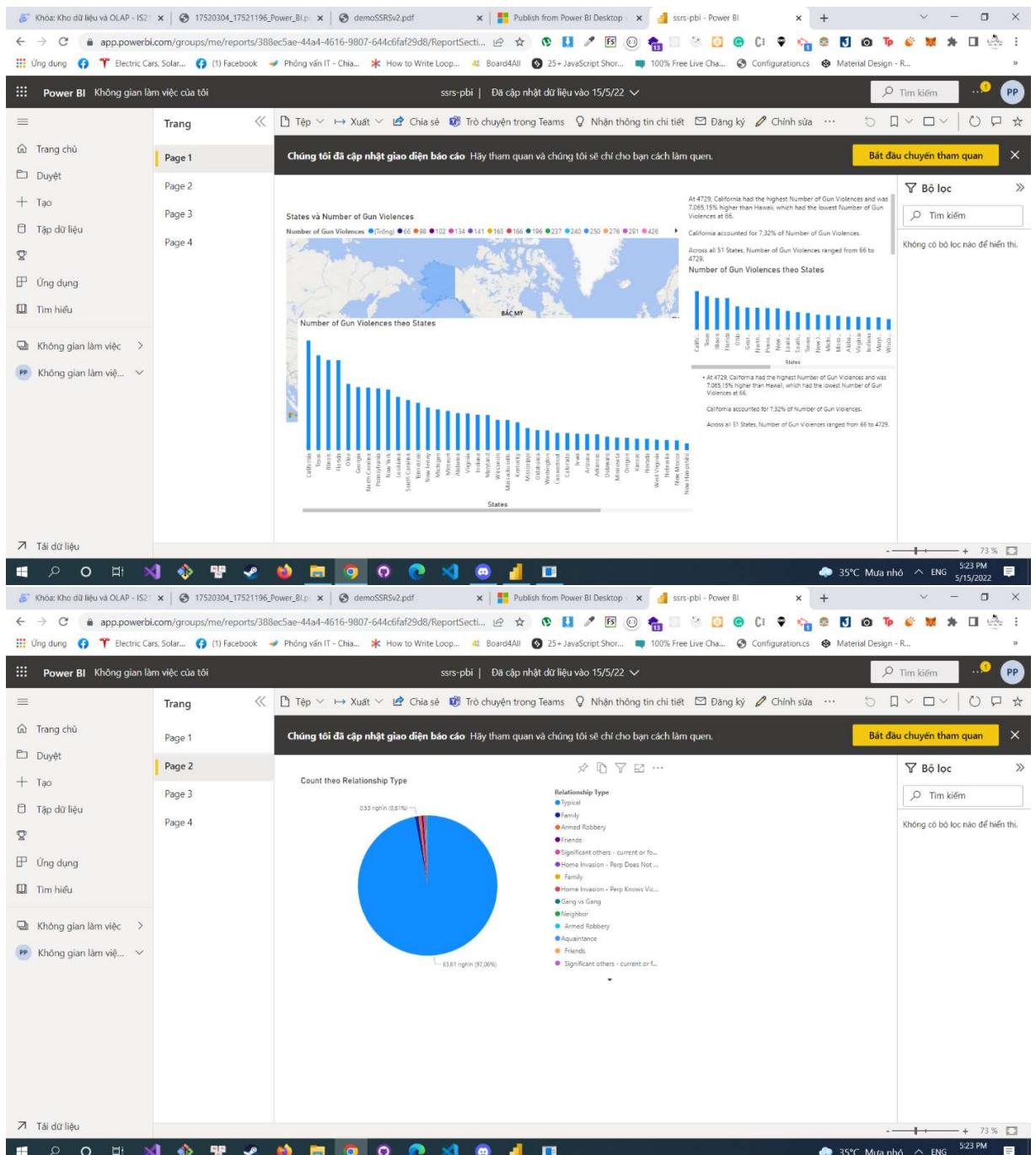
Đợi 1 chút để PowerBI publish



Sau đó ta click vào open... để mở report của chúng ta trên service PowerBI

The screenshot shows a dialog box titled "Publishing to Power BI". It contains a green checkmark icon followed by the text "Success!". Below this, there are two blue links: "Open 'ssrs-pbi.pbix' in Power BI" and "Get Quick Insights". At the bottom left is a yellow circular icon with a bar chart and mobile phone symbol. To its right, the text "Did you know?" is displayed in bold. Below it, there is a message about creating a portrait view for mobile phones and instructions to select "Mobile Layout" on the View tab. A "Learn more" link is also present. On the far right of the dialog box is a yellow "Got it" button.

Ta nhận thấy các report của ta, cùng dataset dữ liệu đã được publish, và có thể chia sẻ liên kết rất dễ dàng sau này có thay đổi gì cũng chỉ cần bấm publish sẽ tự động hoá hết các quy trình trên



Publish Report lên PowerBI Services

#### 4.1.8 Deploy

Link tài nguyên:

[https://drive.google.com/drive/folders/1EqkAhsUQGzGlZNWH8pj\\_pxtFL1paKrQm?usp=sharing](https://drive.google.com/drive/folders/1EqkAhsUQGzGlZNWH8pj_pxtFL1paKrQm?usp=sharing)

## CHƯƠNG 3: DATA MINING

### 1. Giới thiệu Dataset

Dữ liệu về thông số cầu thủ ở giải bóng đá Premier League mùa giải 2015-2016.

<https://www.kaggle.com/datasets/krishanthbarkav/english-premier-leagueepl-player-statistics>

Các thuộc tính (cột) của kho dữ liệu:

- Name: Tên cầu thủ.
- Position: Vị trí thi đấu.
- Appearances: Số lần ra sân.
- Clean sheets: Số trận dữ sạch lưới.
- Goals conceded: Số bàn thua.
- Tackles: Số pha tắc bóng.
- Tackle success %: Tỉ lệ tắc bóng thành công.
- Last man tackles: Số lần tắc bóng cuối sân
- Blocked shots: Số lần chặn cút sút của đối phương
- Interceptions: Số lần tranh chấp.
- Clearances: Số lần phá bóng.
- Headed Clearance: Số lần phá bóng bằng đầu.
- Clearances off line: Số lần cứu thua trên vạch vôi.
- Recoveries: Số lần đoạt bóng thành công.
- Duels won: Số lần tranh chấp thành công.
- Duels lost: Số lần tranh chấp thất bại.
- Successful 50/50s: Số lần thành công trong các pha bóng 50/50.
- Aerial battles won: Số lần tranh chấp bóng bổng thành công.
- Aerial battles lost: Số lần tranh chấp bóng bổng thất bại.

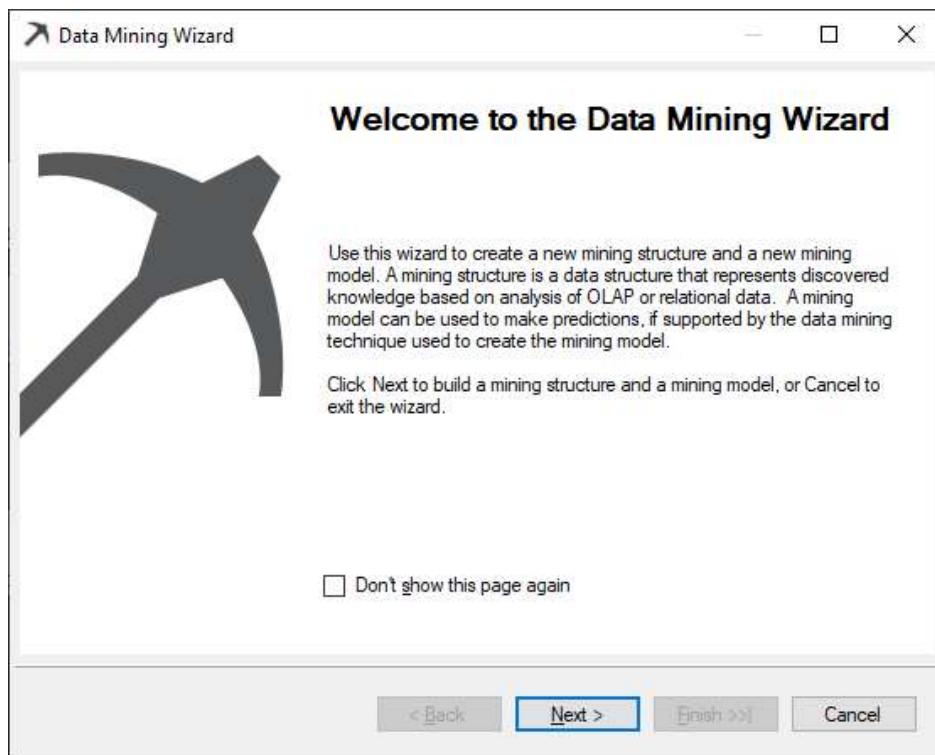
- Own goals: Số bàn phản lưới.
- Errors leading to goal: Số lần mắc sai lầm dẫn tới bàn thua
- Assists: Số kiến tạo.
- Passes: Số đường chuyền.
- Passes per match: Số đường chuyền trung bình mỗi trận.
- Big chances created: Số cơ hội lớn tạo ra.
- Crosses: Số quả tạt
- Cross accuracy %: Tỉ lệ chính xác của các quả tạt
- Through balls: Số đường chuyền chọc khe.
- Accurate long balls: Số đường chuyền dài chính xác.
- Yellow cards: Số thẻ vàng.
- Red cards: Số thẻ đỏ
- Fouls Offsides: Số lần mắc lỗi việt vị.
- Goals: Số bàn thắng.
- Headed goals: Số bàn thắng bằng đầu
- Goals with right foot: Số bàn thắng bằng chân phải.
- Goals with left foot: Số bàn thắng bằng chân trái.
- Hit woodwork: Số lần sút trúng khung gỗ.
- Goals per match: Tỉ lệ bàn thắng mỗi trận.
- Penalties scored: Số bàn thắng ghi được từ penalty.
- Freekicks scored: Số bàn thắng ghi được từ đá phạt.
- Shots: Số cú sút.
- Shots on target: Số cú sút trúng đích.
- Shooting accuracy %: Tỉ lệ sút trúng đích.
- Big chances missed: Số cơ hội mười mươi bị bỏ lỡ.
- Saves: Số pha cứu thua.

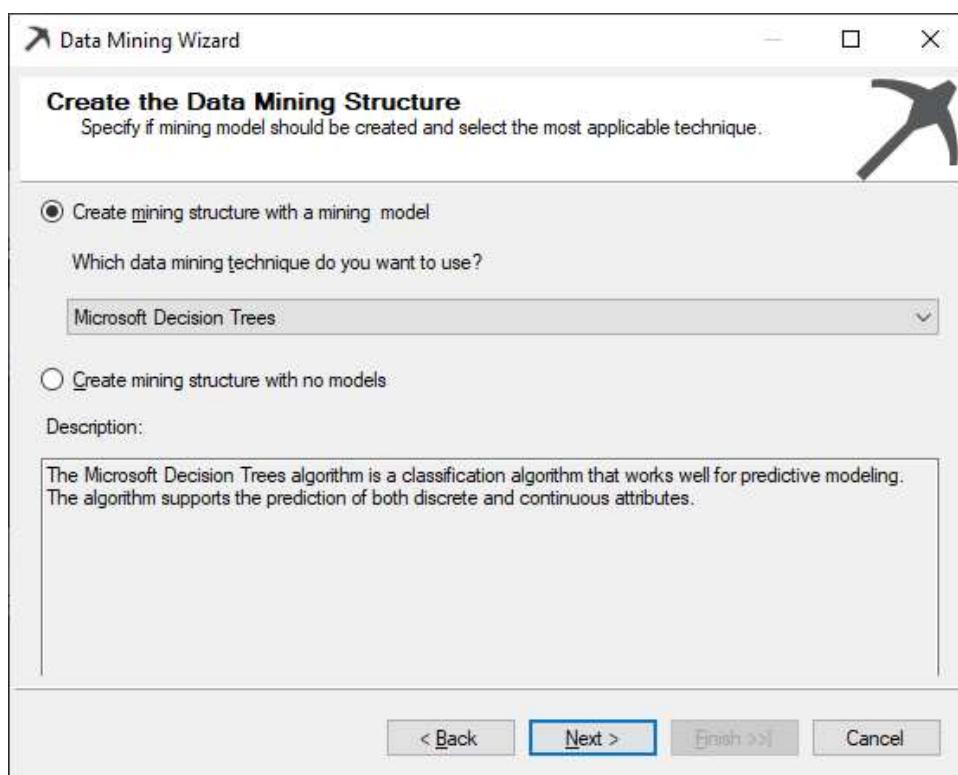
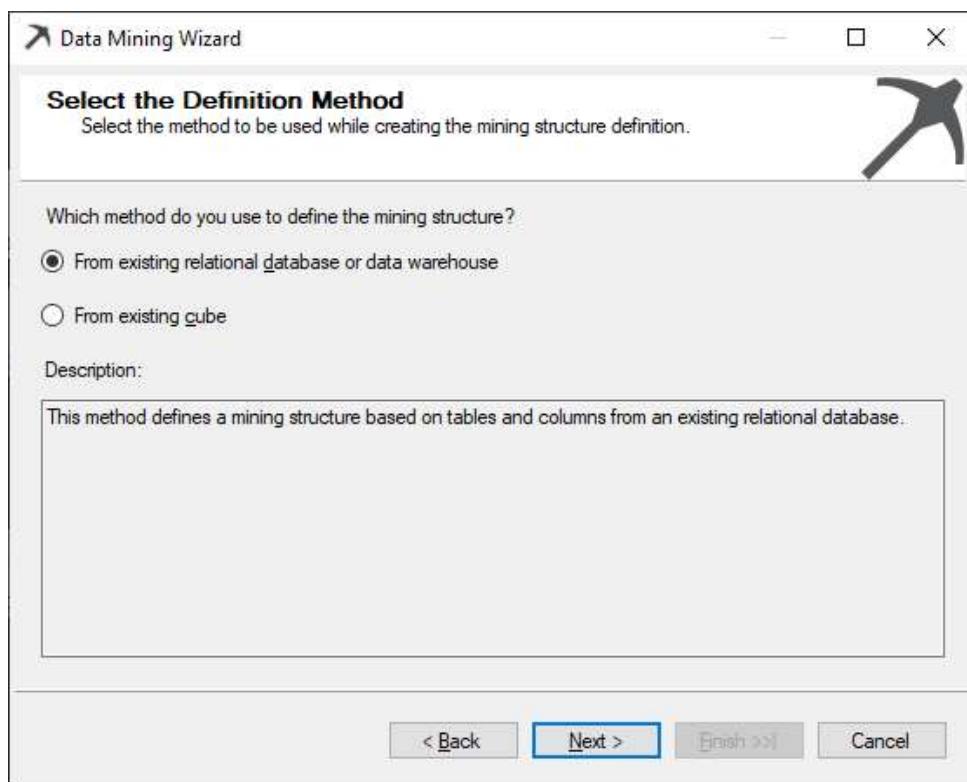
- Penalties saved: Số pha cứu thua penalty.
- Punches: Số pha đấm bóng.
- High Claims: Số pha bắt bóng trên không.
- Catches: Số pha bắt bóng
- Sweeper clearances: Số pha quét bóng.
- Throw outs: Số pha ném bóng.
- Goal Kicks: Số lần thực hiện goal kick.

## 2. Quá trình Data Mining

2.1 Import dữ liệu (GIÓNG SSAS)

2.2 Tạo mô hình Mining





**Select Data Source View**  
Select the data source view to provide the data for the mining structure.

Available data source views:

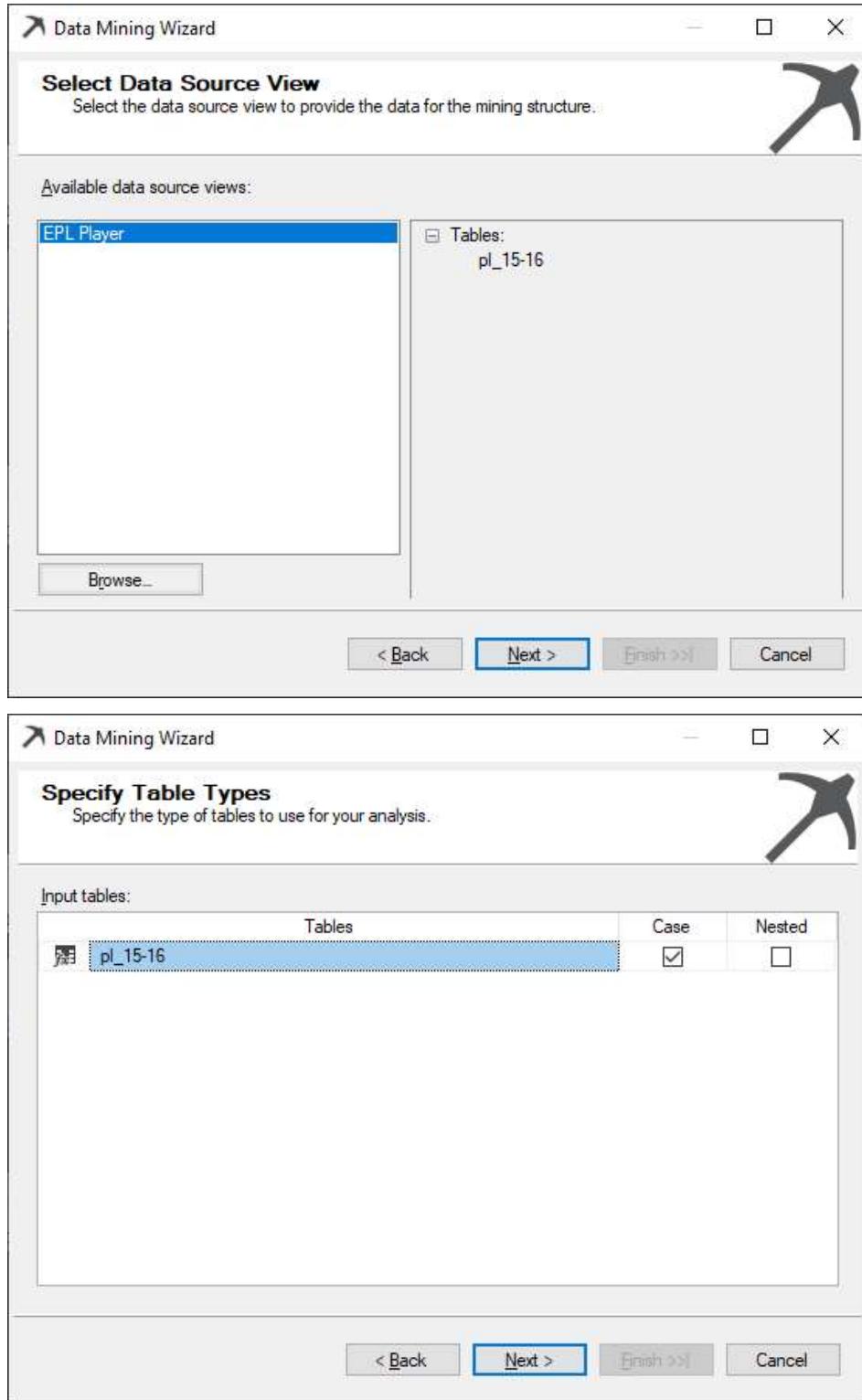
EPL Player	Tables: pl_15-16
------------	---------------------

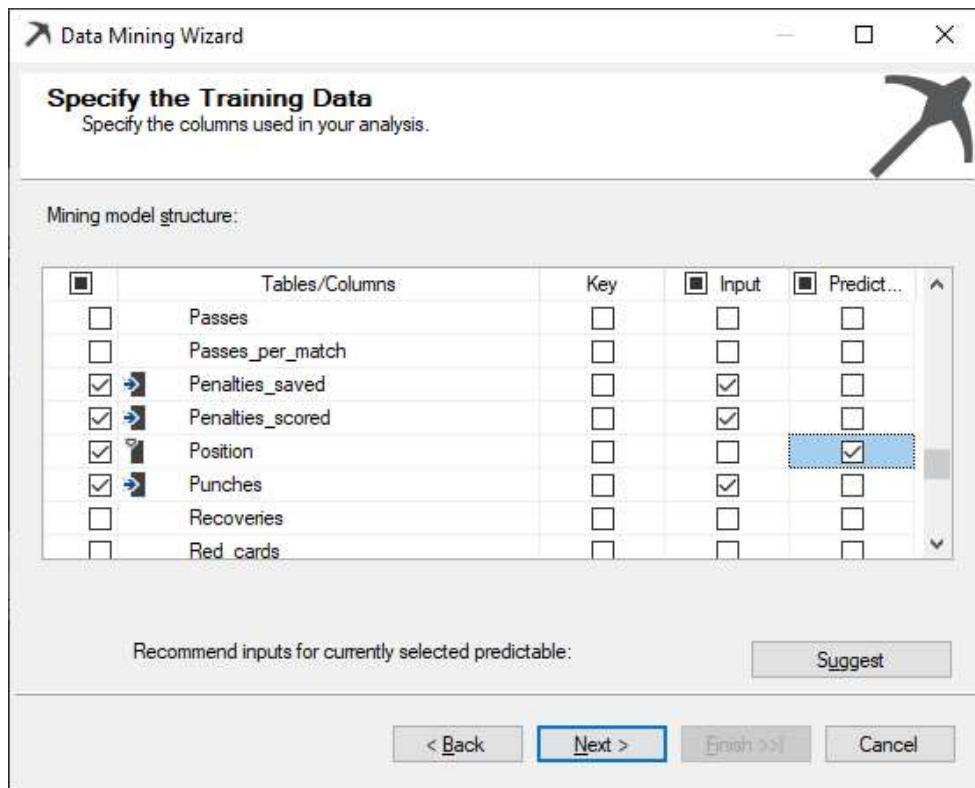
**Specify Table Types**  
Specify the type of tables to use for your analysis.

Input tables:

Tables	Case	Nested
pl_15-16	<input checked="" type="checkbox"/>	<input type="checkbox"/>

< Back    Next >    Finish >>    Cancel

The image shows two windows of the Data Mining Wizard. The top window, titled 'Select Data Source View', displays a list of available data source views, with 'EPL Player' selected. It also shows a table named 'pl\_15-16'. The bottom window, titled 'Specify Table Types', shows the selected table 'pl\_15-16' in a grid with columns for 'Tables', 'Case', and 'Nested'. A checkbox under 'Case' is checked. Both windows have standard Windows-style toolbars at the bottom.



**Data Mining Wizard**

**Specify Columns' Content and Data Type**  
Specify mining structure columns' content and data type.

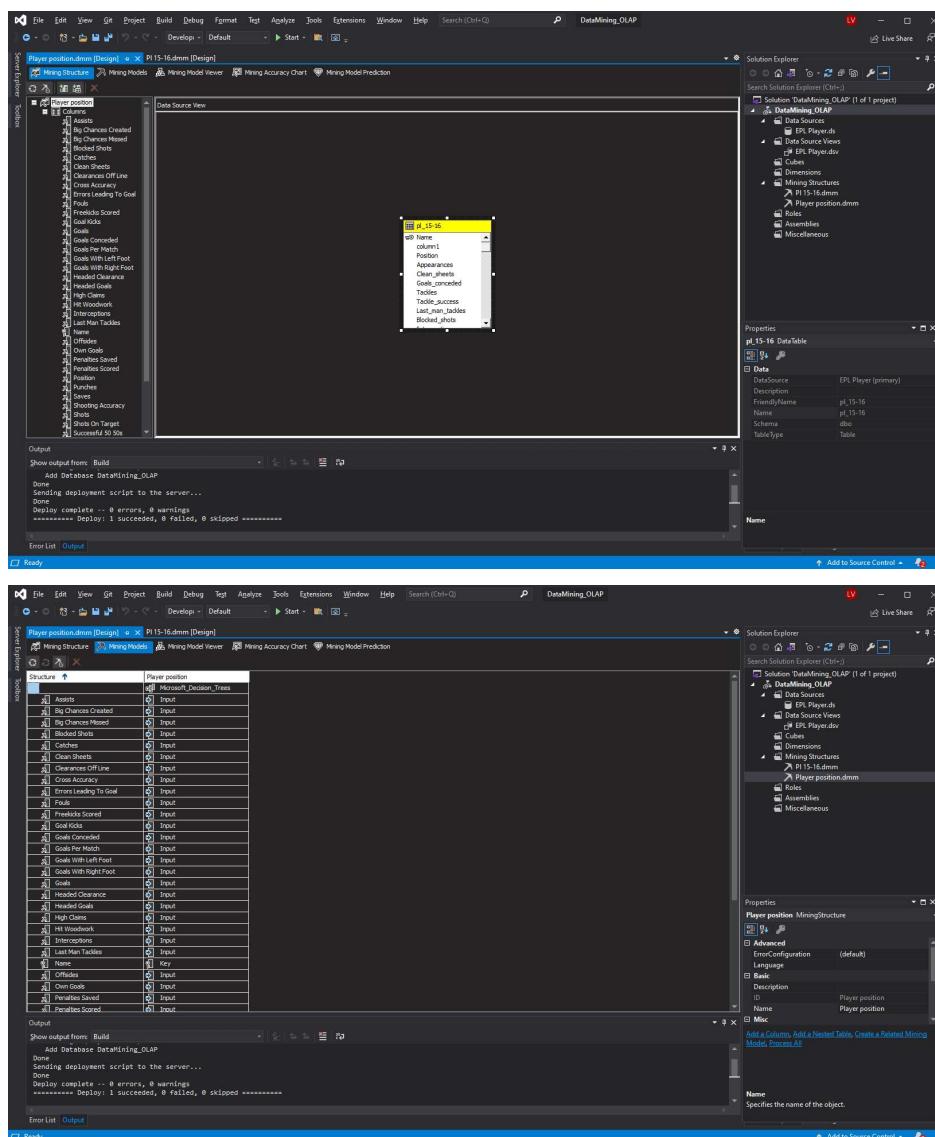
Mining model structure:

Columns	Content Type	Data Type
Aerial Battles Lost	Continuous	Long
Assists	Continuous	Long
Big Chances Created	Continuous	Long
Big Chances Missed	Continuous	Long
Blocked Shots	Continuous	Long
Catches	Discrete	Long
Clean Sheets	Continuous	Long
Clearances Off Line	Discrete	Long
Cross Accuracy	Discrete	Text
Errors Leading To Goal	Continuous	Long
Fouls	Continuous	Long
Freekicks Scored	Discrete	Long
Goal Kicks	Discrete	Long
Goals	Continuous	Long
Goals Conceded	Discrete	Long
Goals Per Match	Continuous	Double
Goals With Left Foot	Continuous	Long
Goals With Right Foot	Continuous	Long
Headed Clearance	Continuous	Long
Headed Goals	Continuous	Long
High Claims	Discrete	Long

Detect continuous or discrete for numeric columns:

## KHO DỮ LIỆU VÀ OLAP

IS217.M22.HTCL



Suggest Related Columns

Column Name	Score	Input
Goals_per_match	0.572	x
Shots	0.571	x
Shooting_accuracy	0.566	x
Goals_conceded	0.545	x
Aerial_battles_won	0.541	
Shots_on_target	0.540	x
Big_chances_missed	0.540	x
Penalties_scored	0.522	x
Clean_sheets	0.522	x
Own_goals	0.515	x
Freekicks_scored	0.515	x
Successful_50_50s	0.511	x
Through_balls	0.502	x
Aerial_battles_lost	0.500	x
Cross_accuracy	0.493	x
Tackle_success	0.486	x
Clearances_off_line	0.474	x
Last_man_tackles	0.474	x
Errors_leading_to_goal	0.407	x
Interceptions	0.396	x
Tackles	0.385	x
Headed_Clearance	0.379	x
Offsides	0.338	x
Blocked_shots	0.330	x
Goals_with_right_foot	0.313	x
Big_chances_created	0.280	x
Headed_goals	0.279	x
Hit_woodwork	0.272	x
Goals_with_left_foot	0.271	x
Sweeper_clearances	0.251	x
Goal_Kicks	0.251	x
Throw_outs	0.251	x
High_Claims	0.251	x
Penalties_saved	0.251	x
Catches	0.251	x
Punches	0.251	x
Saves	0.251	x
Fouls	0.127	x
Goals	0.093	x
Appearances	0.077	
Assists	0.060	x
Yellow_cards	0.039	
Red_cards	0.007	
Accurate_long_balls		
Crosses		
Passes_per_match		
Passes		
Duels_lost		

OK Cancel

Screenshot of the Microsoft Data Mining Modeler interface showing the deployment and processing of a mining structure.

**Deployment Progress:**

- Start time: 5/22/2022 2:05:49 PM
- End time: 5/22/2022 2:05:49 PM
- Duration: 0:00:00
- Process succeeded.

**Processing Progress:**

- Processing Dimension 'Player position->MC-Name' completed.
- Processing Dimension Attribute '(All)' completed.
- Processing Dimension Attribute 'Catches' completed. 19 rows have been read.
- Processing Dimension Attribute 'Clearances Off Line' completed. 5 rows have been read.
- Processing Dimension Attribute 'Cross Accuracy' completed. 40 rows have been read.
- Processing Dimension Attribute 'Freeskicks Scored' completed. 5 rows have been read.
- Processing Dimension Attribute 'Goal Kicks' completed. 42 rows have been read.
- Processing Dimension Attribute 'Goals Conceded' completed. 60 rows have been read.
- Processing Dimension Attribute 'High Claims' completed. 32 rows have been read.
- Processing Dimension Attribute 'Last Man Tackles' completed. 6 rows have been read.
- Processing Dimension Attribute 'Name' completed. 758 rows have been read.
- Processing Dimension Attribute 'Own Goals' completed. 5 rows have been read.
- Processing Dimension Attribute 'Penalties Saved' completed. 6 rows have been read.
- Processing Dimension Attribute 'Penalties Scored' completed. 9 rows have been read.
- Processing Dimension Attribute 'Position' completed. 5 rows have been read.
- Processing Dimension Attribute 'Punches' completed. 22 rows have been read.
- Processing Dimension Attribute 'Saves' completed. 41 rows have been read.
- Processing Dimension Attribute 'Shooting Accuracy' completed. 56 rows have been read.

**Object Model:**

- File Name: Player position.dmm
- Full Path: C:\Users\LLong\source\repos\OLAP\Player position.dmm
- Object ID: Player position
- Object Name: Player position

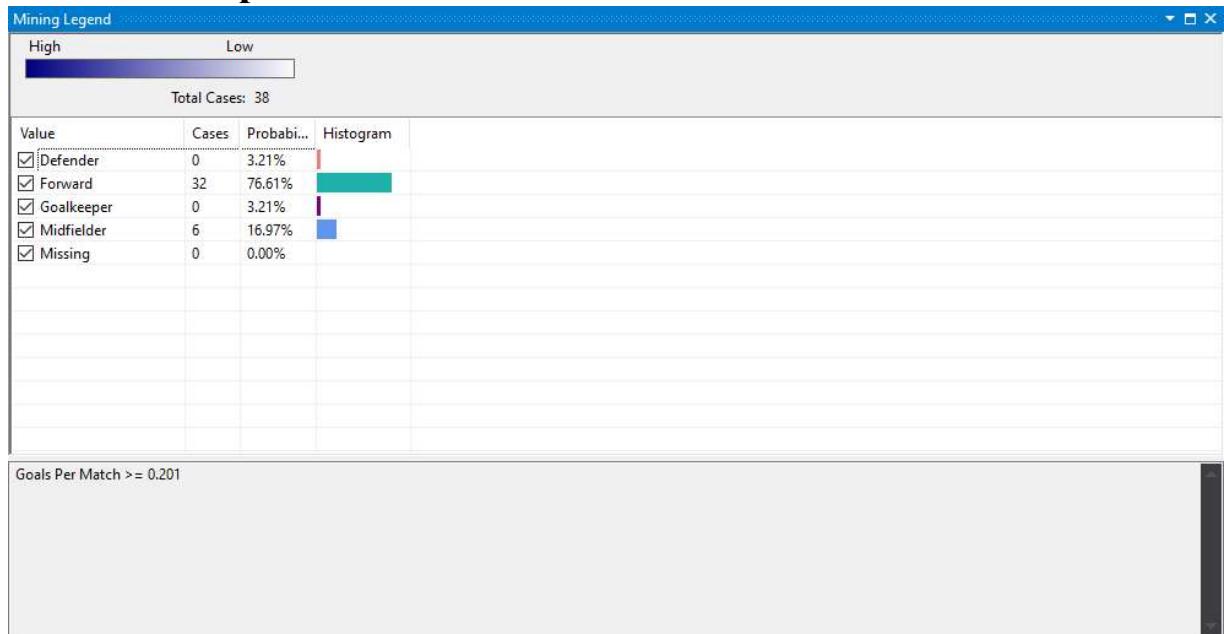
**Decision Tree Diagram:**

```

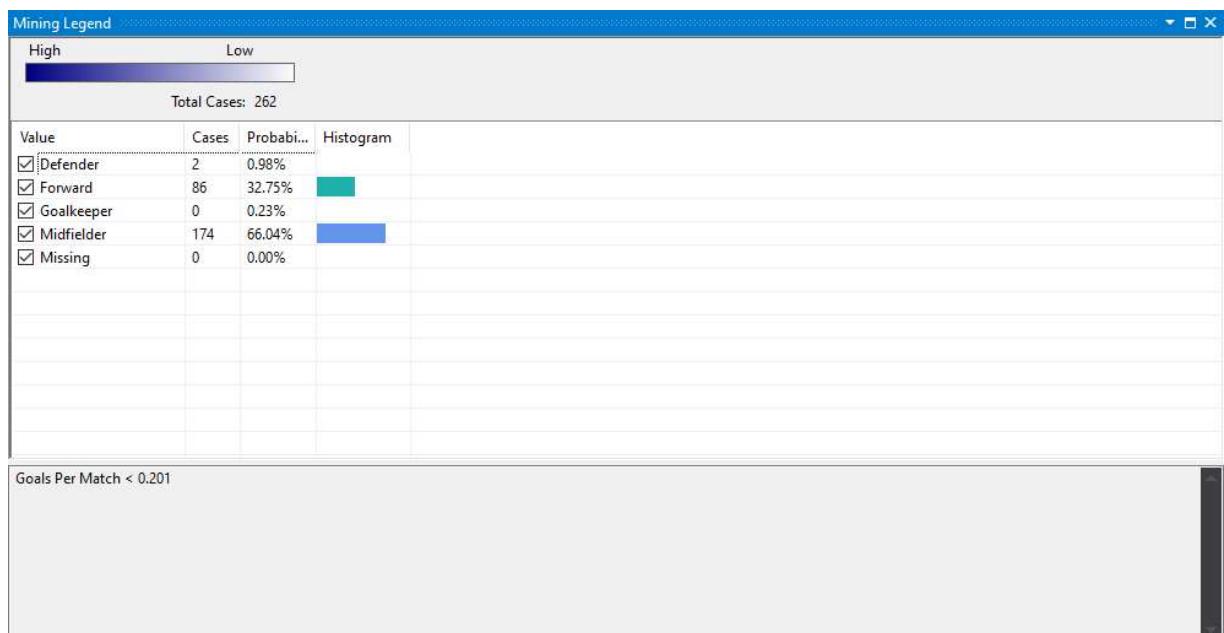
graph TD
    All[All] --> GPM1[Goals Per Match >= 0.201]
    All --> GPM2[Goals Per Match < 0.201]
    All --> GPM3[Goals Per Match = Missing]

    GPM1 --> ABL1[Aerial Battles Lost = Missing]
    GPM1 --> GKN1[Goal Kicks = Missing]
    GPM2 --> ABL2[Aerial Battles Lost not = Missing]
    GPM3 --> GKN2[Goal Kicks not = Missing]
  
```

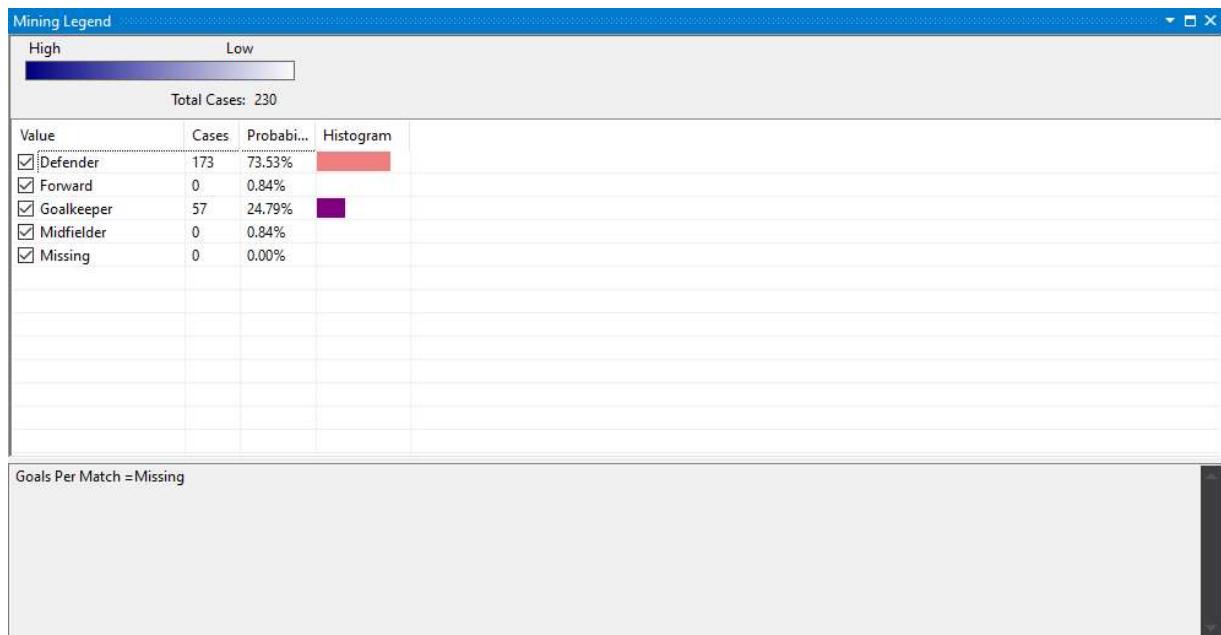
### 3. Tiến hành phân tích.



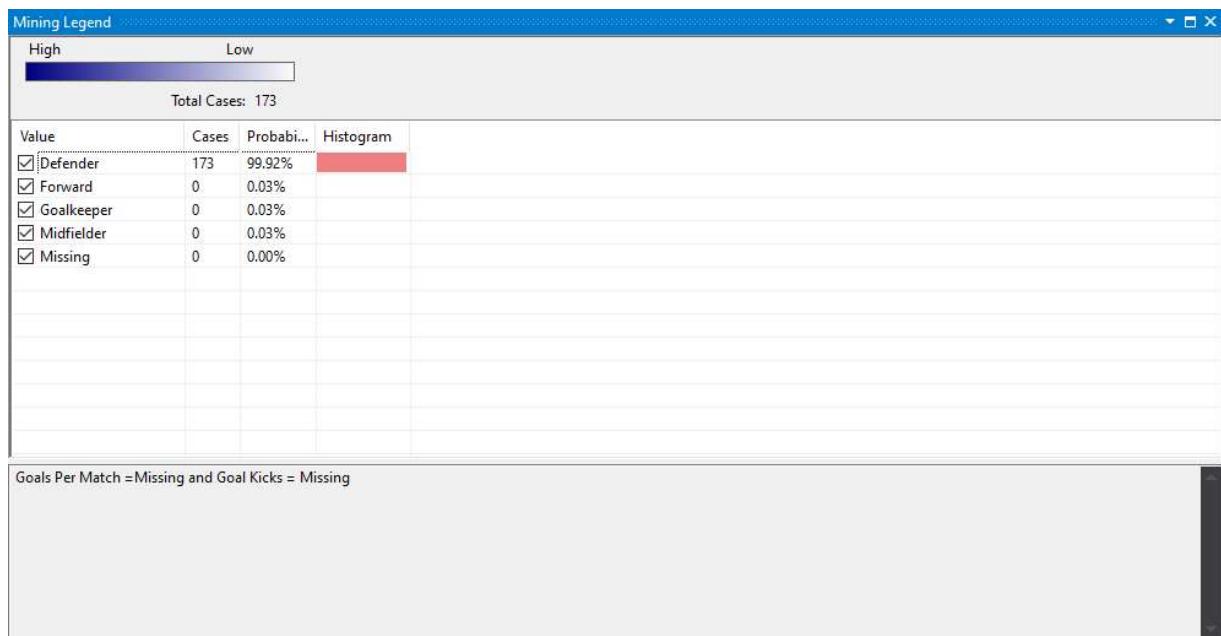
Số bàn thắng mỗi trận lớn hơn hoặc bằng 0.201 thì 76.61% cầu thủ đó là Forward (tiền đạo).



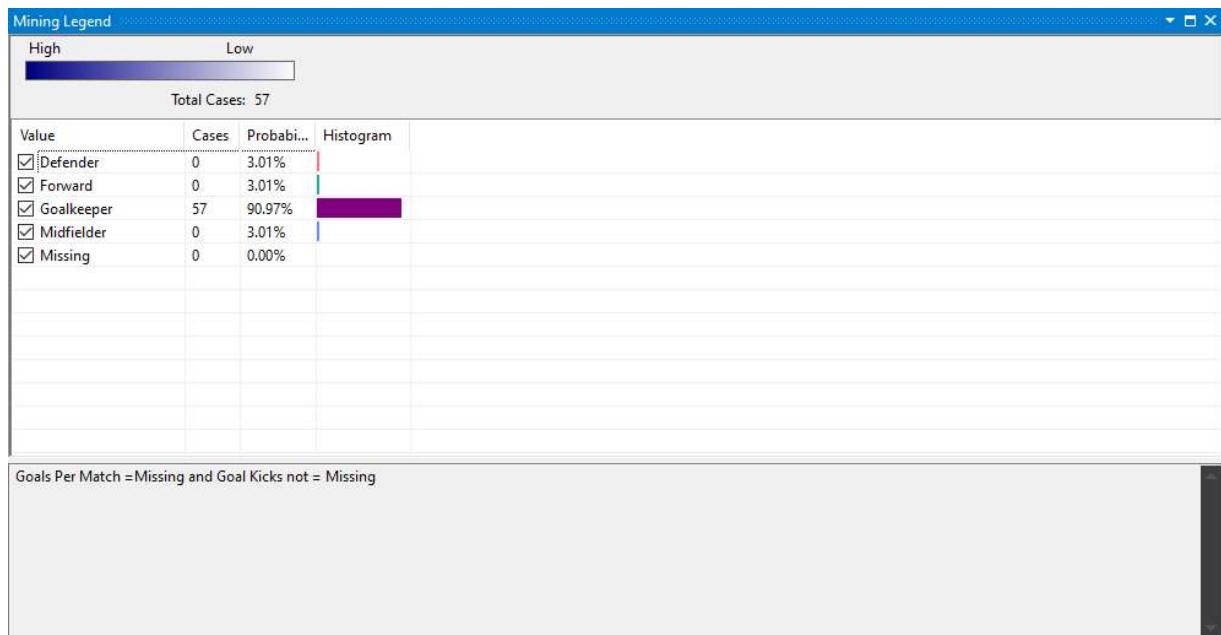
Số bàn thắng mỗi trận nhỏ hơn 0.201 thì 66.04% cầu thủ đó là Midfielder (tiền vệ) và 32.75% cầu thủ đó là Forward (tiền đạo).



Số bàn thắng mỗi trận không tồn tại thì 73.53% cầu thủ đó là Defender (Hậu vệ) và 24.79% cầu thủ đó là thủ môn).



Nếu cả cột bàn thắng mỗi trận và goalkick đều trống thì 99.92% khả năng cầu thủ đó là Hậu vệ.



Nếu cả cột bàn thắng mỗi trận trống và goalkick có tồn tại thì 90.97% khả năng cầu thủ đó là thủ môn.

## CHƯƠNG 4: TÀI LIỆU THAM KHẢO

<https://www.youtube.com/channel/UCd-GLRA2hvJuJFH4TraeyQ>

[https://www.youtube.com/channel/UC\\_zUB0Z8dGcaZfb8HzqLjuQ](https://www.youtube.com/channel/UC_zUB0Z8dGcaZfb8HzqLjuQ)

<https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2022>

<https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/>

## **LINK TÀI NGUYÊN PROJECT**

[https://drive.google.com/drive/folders/19aFyDyDAdAuJ\\_FwoBvFE7ey5TR\\_m0uIk?usp=sharing](https://drive.google.com/drive/folders/19aFyDyDAdAuJ_FwoBvFE7ey5TR_m0uIk?usp=sharing)